Federal Aviation
Administration

International Journal of Applied Aviation Studies

A Publication of the FAA Academy Oklahoma City, Oklahoma

REVIEW PROCESS

The Federal Aviation Administration Academy provides traceability and oversight for each step of the International Journal of Applied Aviation Studies (IJAAS). IJAAS is a peer-reviewed publication, enlisting the support of an international panel of consulting editors. Each consulting editor was chosen for his or her expertise in one or more areas of interest in aviation. Using the blind-review process, three or more consulting editors are selected to appraise each article, judging whether or not it meets the requirements of this publication. In addition to an overall appraisal, a Likert scale is used to measure attitudes regarding individual segments of each article. Articles that are accepted are those that were approved by a majority of judges. Articles that do not meet IJAAS requirements for publication are released back to their author or authors.

Individuals wishing to obtain a copy of the IJAAS on CD may contact Kay Chisholm by email at kay.chisholm@faa.gov, by telephone at (405) 954-3264, or by writing to the following address:

International Journal of Applied Aviation Studies
Kay Chisholm
AMA-800
PO Box 25082
Oklahoma City, OK 73125

POLICY AND DISCLAIMERS

Policy Statement: The Federal Aviation Administration (FAA) Academy strongly supports academic freedom and a researcher's right to publish; therefore, the Federal Aviation Administration Academy as an institution does not endorse the viewpoint or guarantee the technical correctness of any of the articles in this journal.

Disclaimer of Liability: With respect to articles available in this journal, neither the United States Government nor the Federal Aviation Administration Academy nor any of their employees, makes any warranty, express or implied, including the warranties of merchantability and fitness for a particular purpose, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Disclaimer of Endorsement: Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the United States Government or the Federal Aviation Administration Academy. The views and opinions of authors expressed herein do not state or reflect those of the United States Government or the Federal Aviation Administration, and shall not be used for advertising or product endorsement purposes.

PUBLISHER

Academy Superintendent	Gary Condley
Deputy Superintendent	Sunny Lee-Fanning
Executive Editor	Samuel B. Hendrix III

PUBLICATION COMMITTEE

Manager, Regulatory Standards Division	David Long
Manager, Airway Facilities Division	Jessie McMullen
Manager, Air Traffic Division	David Haddad
Manager, Airports & International Training Division	Sherry Reese

EDITORIAL STAFF

Managing Editor	Kay Chisholm
Production and Marketing Editor	Laurie Wilson
Associate Editor	Deann King
Research Consultant	Todd P. Hubbard

CONSULTING EDITORS

International

Ruwantissa I.R. Abeyratne	ICAO, Montreal, Canada
Pierre R. Band	Institut du Cancer de Montréal, Québec, Canada
Simon Bennett	Institute of Lifelong Learning, Leicester, UK
Robert Bor	Royal Free Hospital, London, UK
Peter Brooker	Cranfield University, Cranfield, UK
Bryan Burke	Macquaire University, Sydney, Australia
Finian Connolly	Executive & Professional Training Institute, Drogheda, Co Louth, Ireland
Hugh David	Eurocontrol, East Sussex, UK
Sidney Dekker	Linköping Institute of Technology, Linköping, Sweden
Teresa C. D'Oliveira	ISPA, Lisbon, Portugal
Robert van Doorn	Universiteit Maastricht, The Netherlands
Alois Farthofer	Human Factors Research, St. Georgen, Austria
Massimo Felici	University of Edinburgh, Edinburgh, UK
Rodney Fewings	Cranfield University, Cranfield, Bedfordshire UK
Gerard J. Fogarty	University of Southern Queensland, Toowoomba, Australia
Jarle Gimmestad	Braathens Airlines, Fornebu, Norway
Gael P. Hammer	Johannes Gutenberg-University, Mainz, Germany
Don Harris	Cranfield University, Cranfield, Bedfordshire, UK
Irene Henley	University of Western Sidney, Sidney, Australia
Graham Hunt	Massey University, Auckland, New Zealand
Magnus Jacobssen	Linköping Institute of Technology, Linköping, Sweden
Lynn M. Jeffrey	Massey University Auckland, Auckland, New Zealand
K. Wolfgang Kallus	Karl Franzens University, Graz, Austria
Michael John Kay	RMIT University, Melbourne, Victoria, Australia
Elizabeth Mathews	Aviation English Services, Wellington , New Zealand
Jeremy Mell	French Embassy, Washington DC
Jim Mitchell	University of Western Sydney, Penrith South, Australia

Kjell Mjøs   Norwegian University of Science and Technology,
                              Oslo, Norway
Brett Molesworth   University of New South Wales, Sydney, Australia
Jan Noyes   University of Bristol, Bristol, UK
David O'Hare   University of Otago, Dunedin, New Zealand
Esther Oprins   Air Traffic Control the Netherlands, Schiphol Airport,
                              Netherlands
Jens Rolfsen   Det Norske Veritas, Høvik, Norway
Paul Roosens   University of Antwerp, Antwerpen, Belgium
Michael Siegrist   Institute for Environmental Decisions, Zurich, Switzerland
T. Leigh Signal   Massey University, Wellington, New Zealand
M. "Mattie" Tops   Leiden University, Leiden, The Netherlands
Steve Thatcher   University of South Australia, Mawson Lakes, Australia
Claudia van de Wal   VALK Foundation, Netherlands
Lucas van Gerwen   VALK Foundation, Netherlands
Leo Vermeulen   University of Pretoria, Pretoria, South Africa
Stephen Walsh   Interaction Training Associates, Netherlands
Damien J. Williams   University of Bristol, Bristol, UK
Robert Wolfger   Austrian Airlines, Strasshof, Austria
Rogier Woltjer   Linköping Institute of Technology, Linköping, Sweden
Beat Zimmermann   Air Navigation Institute, Thun, Switzerland

U.S.A.

Robert I. Aceves   St. Cloud State University, St. Cloud, MN
Vicki Ahlstrom   FAA, William J. Hughes Technical Center, Atlantic City, NJ
Amy Alexander   Aptima, Inc., Woburn, MA
Steve Anderson   St. Cloud State University, St. Cloud, MN
Francis "Frank" Ayers   Embry-Riddle Aeronautical University, Daytona Beach, FL
Larry Bailey   FAA Academy, Oklahoma City, OK
R. Kurt Barnhart   Kansas State University, Salina, KS
Robert Baron   The Aviation Consulting Group, Myrtle Beach, SC
Ellen Bass   University of Virginia, Charlottesville, VA
J. Matthew Beaubien   American Institutes for Research, Washington, DC
Wendy Beckman   Parks College, St. Louis University, St. Louis, MO
Dennis B Beringer   Civil Aerospace Medical Institute, Oklahoma City, OK
Robert O. Besco   Professional Performance Improvement, Dallas, TX
Aleta Best   Department of Transportation, Washington, DC
Kate Bleckley   Civil Aerospace Medical Institute, Oklahoma City, OK
James P. Bliss   Old Dominion University, Norfolk, VA
Deborah Boehm-Davis   George Mason University, Fairfax, VA
Brent Bowen   University of Nebraska at Omaha, Omaha, NE
Jonathan Bricker   University of Washington, Seattle, WA
Christopher R. Brinton   Mosaic ATM, INC., Leesburg, VA
Dana Broach   Civil Aerospace Medical Institute, Oklahoma City, OK
Philip Budd   Graduate Studies in Counseling, Bethany, OK
Ernesto A. Bustamante   University of Idaho, Moscow, ID
James N Butcher   University of Minnesota, Minneapolis, MN
Evan Byrne   National Transportation Safety Board, Washington, DC

John Cain	Florida Technology Institute, Melbourne, FL
Raymond E. Cain, Jr.	Florida Memorial University, Miami Gardens, FL
John A. Caldwell	Wright-Patterson Air Force Base, Dayton, OH
Kim Cardosi	Volpe National Transportation Systems Center, Cambridge, MA
Stephen M. Casner	NASA Ames Research Center, Moffett Field, CA
Thomas R. Chidester	Civil Aerospace Medical Institute, Oklahoma City, OK
Robert H. Chisholm	Support Systems Associates, Melbourne, FL
Maria Consiglio	NASA Langley Research Center, Hampton, VA
David Conway	Southeastern Oklahoma State University, Durant, OK
Paul A. Craig	Middle Tennessee State University, Murfreesboro, TN
Jerry Crutchfield	Civil Aerospace Medical Institute, Oklahoma City, OK
John Deaton	Florida Institute of Technology, Melbourne, FL
Michael DeGrave	Kenyon & Kenyon LLP, New York, NY
Carmen Delgado-Morales	FAA Academy, Oklahoma City, OK
Charles DeJohn	Civil Aerospace Medical Institute, Oklahoma City, OK
James DeVoll	Office of Aerospace Medicine, Washington, DC
Archie Dillard	FAA National Headquarters, Washington, DC
Brian G. Dillman	Purdue University, West Lafayette, IN
Key Dismukes	NASA Ames Research Center, Moffett Field, CA
Nancy Dorighi	Ames Research Center, Moffett Field, CA
Wayne Dornan	Middle Tennessee State University, Murfreesboro, TN
Graham Elliott	FAA Academy, Oklahoma City, OK
Boyd Falconer	Russell Reynolds Associates, Chicago, IL
Richard Fanjoy	Purdue University, West Lafayette, IN
Douglas R. Farrow	Federal Aviation Administration, Washington, DC
Wai-Tat Fu	University of Illinois, Savoy, IL
Kenneth H. Funk II	Oregon State University, Corvallis, OR
Peter C. Gardiner	SCSI, Torrance, CA
Kenneth Gardner	FAA, Flight Standards, New Cumberland, PA
Patricia Garrambone	FAA, Boston TRACON, Boston, MA
A. F. Grandt	Purdue University, West Lafayette, IN
Mavis Green	Harvard University, Cambridge, MA
Carla A. Hackworth	Civil Aerospace Medical Institute, Oklahoma City, OK
Chris Hallman	Great Circle Consulting, Inc., Newman, GA
Steven Hampton	Embry-Riddle Aeronautical University, Daytona Beach, FL
Jeffrey T. Hansberger	George Mason University, Fairfax, VA
Frederick D. Hansen	Oklahoma State University - Tulsa, Tulsa, OK
Samuel B. Hendrix III	FAA Academy, Oklahoma City, OK
Kevin High	Western Michigan University, Battle Creek, MI
Jerry Higley	FAA Academy, Oklahoma City, OK
Alan Hobbs	NASA Ames Research Center, Moffett Field, CA
Judy Holcomb	FAA Academy, Oklahoma City, OK
Kent Holtorf	Holtorf Medical Group, Torrance, CA
Willem Homan	Western Michigan University, Battle Creek, MI
Amy Hoover	Central Washington University, Ellensburg, WA
Todd P. Hubbard	Oklahoma State University, Stillwater, OK
David R. Hunter	Artis, LLC, Reston, VA
Peter Hwoschinsky	FAA National Headquarters, Washington, DC
Sig Illing	FAA Academy, Oklahoma City, OK

Kathi Ishimaru	FAA, Northwest Mountain Regional Office, Renton, WA
Florian Jentsch	University of Central Florida, Orlando, FL
Jeffrey Johnson	St. Cloud State University, St. Cloud, MN
Merrill R. Karp	Arizona State University, Mesa, AZ
Deann King	FAA Academy, Oklahoma City, OK
Raymond E. King	Civil Aerospace Medical Institute, Oklahoma City, OK
William Knecht	Civil Aerospace Medical Institute, Oklahoma City, OK
Jefferson M. Koonce	Technology Systems, Inc. Wiscasset, ME
Bruce Kuhlmann	FAA Academy, Oklahoma City, OK
Mary N. Kutz	Oklahoma State University, Stillwater, OK
Florian Jentsch	University of Central Florida, Orlando, FL
Victor LaSaxon	The Boeing Company, Midwest City, OK
Chien-tsung Lu	Central Missouri State University, Warrensburg, MO
Rebecca Lutte	University of Nebraska at Omaha, Omaha, NE
Richard L. Mangrum	Kent State University, Kent, OH
Carol Manning	Civil Aerospace Medical Institute, Oklahoma City, OK
Steven Marks	Oklahoma State University, Stillwater, OK
Royce Ann Martin	Bowling Green State University, Bowling Green, OH
Celeste Mayer	North Carolina State University, Raleigh, NC
Raymon M. McAdaragh	FAA, Hampton, VA
H. C. McClure	MAC Air Consulting, Queen Creek, AZ
Kathleen L. McFadden	Northern Illinois University, DeKalb, IL
Jeremy Mell	French Embassy, Washington, DC
Harry Minniear	Indiana State University, Terre Haute, IN
Marjo Mitsutomi	University of Redlands, Redlands, CA
Richard H. Mogford	NASA Ames Research Center, Moffett Field, CA
John Morey	Dynamics Research Corporation, Andover, MA
Kathleen Mosier	San Francisco State University, San Francisco, CA
Vahid Motevalli	The George Washington University, Ashburn, VA
Martin Mumenthaler	Stanford University, Redwood City, CA
Tom Nesthus	Civil Aerospace Medical Institute, Oklahoma City, OK
Michael S. Nolan	Purdue University, West Lafayette, IN
Gary J. Northam	Parks College, St. Louis University, St. Louis, MO
Dale B. Oderman	Purdue University, West Lafayette, IN
Jim Oppermann	Ohio State University, Columbus, OH
Korhan Oyman	Florida Institute of Technology, Melbourne, FL
Allen J. Parmet	Midwest Occupational Medicine, Kansas City, MO
Manoj Patankar	Parks College, St. Louis University, St. Louis, MO
Donald A. Petrin	Purdue University, West Lafayette, IN
James J. Picano	Mind Quest, LLC, Suisun, CA
Jean Potvin	Saint Louis University, St. Louis, MO
Julia Pounds	Civil Aerospace Medical Institute, Oklahoma City, OK
Roni Prinzo	Civil Aerospace Medical Institute, Oklahoma City, OK
Edward Pugacz	FAA, William J. Hughes Technical Center, Atlantic City, NJ
Yandong Qiang	Johns Hopkins University, Baltimore, MD
Stephen M. Quilty	Bowling Green State University, Bowling Green, OH
William B. Rankin II	Florida Memorial College, Miami, FL

Esa M. Rantanen   Rochester Institute of Technology, Rochester, NY
Robert Ripley   Auburn University, Auburn, AL
Vladimir Risukhin   Western Michigan University, Kalamazoo, MI
Charles L. Robertson   University of North Dakota, Grand Forks, ND
Vladimir Risukhin   Western Michigan University, Kalamazoo, MI
Jason J. Saleem   Indiana Univ.-Purdue University, Indianapolis IN
Steven M. Samuels   US Air Force Academy, Colorado Springs, CO
Nadine Sarter   The Ohio State University, Columbus, OH
David Schroeder   Civil Aerospace Medical Institute, Oklahoma City, OK
Gregory L. Schwab   Indiana State University, Terre Haute, IN
Scott Shappell   Clemson University, Clemson, SC
J. Anthony Sharp   Elizabeth City State University, Elizabeth City, NC
John W. Sheremeta, Jr.   American Airlines, Lake Ronkonkoma, NY
Mark A. Sherman   Ohio University, Athens, OH
Kenneth Sperry   The Boeing Company, Midwest City, OK
John E. Stewart II   U.S. Army Research Institute, Fort Rucker, AL
Alan J. Stolzer   Parks College, St. Louis University, St. Louis, MO
Judith B. Strother   Florida Institute of Technology, Melbourne, FL
Craig Gerald Stroup   American West Airlines, Phoenix, AZ
Terry von Thaden   University of Illinois, Savoy, IL
Judith Foss Van Zante   NASA Glenn Research Center, Cleveland, OH
Stephen Véronneau   CAMI, Oklahoma City, OK
Ken Wallston   Vanderbilt University, Nashville, TN
Ron Ward   FAA Academy, Oklahoma City, OK
Thomas Weitzel   Embry-Riddle Aeronautical University, Daytona Beach, FL
Michael Wetmore   Northcentral University, Prescott Valley, AZ
Christopher Wickens   University of Illinois Urbana-Champaign, Urbana, IL
Michael Wiggins   Embry-Riddle Aeronautical University, Daytona Beach, FL
Kevin Williams   Civil Aerospace Medical Institute, Oklahoma City, OK
Dale Wilson   Central Washington University, Ellensburg, WA
Donna Forsyth Wilt   Florida Institute of Technology, Melbourne, FL
Peter Wolfe   Professional Aviation Board of Certification, Washington, DC
John Young   Purdue University, West Lafayette, IN

PHILOSOPHY STATEMENT

Cornelius Lanczos, a mathematician working in the field of applied analysis, expressed the history of mathematics in three phases:

1) A given physical situation is translated into the realm of numbers,
2) By purely formal operations with these numbers certain mathematical results are obtained, [and]
3) These results are translated back into the world of physical reality (1988, p. 1). [1]

Formal papers, in subjects related to aviation, roughly follow the same course. However, there appears to be a weakness in aviation research, that being the omission of the third phase.

It is not good enough that conclusions are drawn, if those conclusions fail to improve the system observed. Clearly, the observed have a say in implementing the conclusions of research, but their failure to implement the conclusions drawn by the researcher may be more indicative of a lack of understanding than a lack of desire. Researchers tend to peer into complex systems as through a soda straw, forming formal opinions on the finite without understanding the complete system. Industry, ever mindful of the complete system, may find research irrelevant because it makes much to do about nothing.

The editorial staff, to include those listed as consulting editors, is committed to the improvement of all individuals within the aviation community. We seek to enhance existing systems bearing in mind that small improvements must not upset the delicate balance between too little and too much help. We also seek to promote safety, not by lip service, but by demonstration in how we execute our studies and how we report our findings.

We feel that the best way to translate results back to the physical world is to incorporate the viewpoints of people around the globe. Without the influence of a worldwide community, we deny the significance of diversity and ignore the perspectives of gifted scientists from different countries. It is our hope that each reader will feel the same.

---

[1]Lanczos, C. (1988). Applied Analysis. Mineola, NY: Dover Publications, Inc.

EDITOR'S NOTES

Our lead article by *Gerard J. Fogarty and Elizabeth Buikstra* reports on a study that tested hypotheses about different causal pathways for errors and procedural violations. *A Test of Direct and Indirect Pathways Linking Safety Climate, Psychological Health, and Unsafe Behaviours* examines the relationship between workplace safety climate factors, individual psychological health factors, and self-reported errors and violation behaviors. Three hundred eight aviation maintenance engineers completed a self-report questionnaire. The study demonstrates the importance of including both organizational and individual level variables to assess the safety status of an organization with expanded safety climate surveys.

Steve Jarvis and Don Harris' research on glider accidents demonstrates the benefit of examining the flight phase in which the seminal event of an accident occurs. *Investigation into Accident Initiation Events by Flight Phase, for Highly Inexperienced Glider Pilots* targets appropriate remedial actions and the use of flying exposure measures to produce comparable accident rate data rather than simply comparing accident counts. All pilot-related accidents in the British Gliding Association database (2002 – 2006) were identified and accident rates were calculated for each flight phase.

Our third paper is the conclusion of a three part series on runway incursions by *William B. Rankin, II. Runway Incursions: An Industry Examination of FAA Initiatives and Objectives* compares the perceptions of industry officials to the FAA's Runway Incursion Plan of 1991 and the Runway Safety Blueprint 2002-2004 to see if there is a continued similarity of the perceived effectiveness of the FAA initiatives or objectives. Since airport driver training was ranked as the number one initiate in the 1994 study and was not included in the FAA Runway Safety Blueprint 2002-2004, the 2007 study asked industry officials if airport driver training should, or should not be included in the FAA Runway Safety Blueprint.

Ernesto A. Bustamante's research demonstrates the superior advantage of using likelihood alarm technology (LAT) to increase decision-making accuracy, decrease decision-making bias, and ultimately enhance monitoring performance. *Implementing Likelihood Alarm Technology in Integrated Aviation Displays for Enhancing Decision-Making: A Two-Stage Signal Detection Modeling Approach* presents a two-stage signal detection modeling approach of decision making while interacting with integrated aviation displays that allows researchers to partition these separate processing stages.

In *Locus of Control and Self-Attribution as Mediators of Hazardous Attitudes among Aviators: A Review and Suggested Applications, John E. Stewart* examines and addresses LOC in the context of hazardous attitudes. This paper examines concepts from attribution theory, and contends that these are consistent with the processes underlying the maintenance of LOC and hazardous attitudes. It is recommended that integration of LOC and attribution theory should provide an enhanced explanation of the motivational bases for risk taking and decision making among aviators.

Beth M. Beaudin-Seiler, Jeffrey M. Beaubien, and *Ryan C. Seiler's* paper *Collegiate Flight Training: Making Progress in the Face of Adverse Conditions* presents the approach a collegiate flight program took to better track the progress of its student pilots. Results show that gaps in training explain significant criterion. Newly developed tools, such as the *Gaps in Instruction Adjustment Matrix,* may help to standardize the administrative decisions concerning the amount of remedial training required following a gap in instruction.

Incorrect maintenance information is a contributing factor in a number of recent aircraft mishaps. *Bonnie Lida Rogers, Christopher J. Hamblin,* and *Alex Chaparro* study the types of errors found in aircraft maintenance manuals published by manufacturers. In *Classification and Analysis of Errors Reported in Aircraft Maintenance Manuals* the authors analyze Publication Change Requests (PCRs) to document the most frequently reported types of errors found in aircraft maintenance manuals, to identify how errors vary across Air Transport Association (ATA) chapters, and identify the corrective actions required to address the cited problem.

Fatigue plays a major role in many aviation accidents and incidents. In *Effects of Fatigue on Flight Training: A Survey of U.S. Part 141 Flight Schools, Sara McDale* and *Jiao Ma* focus on the flight instructor. Due to the traditionally long workday and intensive workload, flight instructors are particularly subject to fatigue. A national survey was conducted to assess Part 141 flight school instructors' self-awareness of their fatigue issues, impact of fatigue on quality and safety, and potential solutions. Instructors reported that fatigue had negatively affected flight instruction.

Elizabeth T. Newlin, Ernesto A. Bustamante, and *James P. Bliss'* study *Alarm Relevance and Reliability: Factors Affecting Alarm Responses by Commercial Pilots,* assesses the influence of alarm relevance and reliability on pilots' perceptions of relevance, urgency, importance, how compelled they were to respond, and actual response behavior. The findings suggest that pilots consider alarm relevance when responding to alarms but they are compelled to respond to unreliable alarms because of their training. Alarm relevance affects pilots' rate and speed of response, and pilots are influenced by their training to overmatch their alarm responses.

In *Stress in Ballooning: An Exploratory Cortisol Study, AJ de Voogt* explores the possible existence of stress in balloon operations by measuring stress-related hormonal changes during balloon flights. Salivary cortisol was measured in experienced balloon-pilots before, during, and after a balloon flight. Though further research is necessary, the data suggest that even in experienced pilots, balloon flights may be stressful and therefore may influence the risk for pilot errors.

The FAA bases its fundamentals of instruction (FOI) primarily on principals of cognitive theory and behaviorism. *Amy L. Hoover's* developmental paper *Educa-*

tional Learning Theories: Informing the Fundamentals of Instruction, addresses the concern that flight instructor applicants may pass the FOI knowledge exam without gaining a complete understanding of important underlying educational learning theories applicable to flight training.   Examples from the educational literature are used to describe some of those social learning theories and relate them to design and delivery of flight training curricula to enhance the transition from theory to practice.

KC

International Journal of Applied Aviation Studies

Volume 8, Number 2   2008

Papers

A Test of Direct and Indirect Pathways Linking Safety Climate, Psychological Health, and Unsafe Behaviours

Gerard J. Fogarty and Elizabeth Buikstra
University of Southern Queensland,
Toowoomba
Australia

Abstract

This study examined the relationship between workplace safety climate factors, individual psychological health factors, and self-reported errors and violation behaviours. The aim was to test hypotheses about different causal pathways for errors and procedural violations. Three hundred eight aviation maintenance engineers completed a self-report questionnaire developed for this study. A structural model depicting how workplace violations and psychological health act as mediators between safety climate and errors was tested using structural equation modelling. The model fitted the data with safety climate accounting for 63% of the variance in violations and 52% of the variance in psychological health. Violations and psychological health combined to predict 58% of the variance in errors. The study demonstrates the importance of including both organizational and individual level variables to assess the safety status of an organization through the use of expanded safety climate surveys.

A Test of Direct and Indirect Pathways Linking Safety Climate,
Psychological Health, and Unsafe Behaviours

William James (1890) first introduced the concept of everyday cognitive failings in the late 1800s, but sustained scientific interest in the concept had to await the advent of complex industrial technologies that stretched workers to their physical and mental limits. High-risk organizations such as offshore oil, nuclear power, chemical processing plants and aviation are unforgiving environments where errors can have devastating consequences. Growing concern about the cause of errors has led researchers to consider the impact of constructs such as safety climate, attitudes, social norms, stress, and cognition on safety behaviours such as errors and violations. However, much of this research is piecemeal. What is needed in the literature are studies that bring together these constructs in structural models that can be tested, thus providing empirical support for what are sometimes no more than descriptive models of error causation. The present study used structural equation modelling to test models of the direct and indirect effects of safety climate factors and individual psychological health on self-reported errors and violations in aviation maintenance. In the sections that follow, we trace

the development of theory in this area of safety research before describing the development of a model that posits different causal pathways for errors and violations.

A group of researchers working mainly in the offshore oil industry (Fleming, Flin, Mearns, & Gordon, 1998; Mearns & Flin, 1999; Mearns, Flin, Gordon, & Fleming, 2001) modelled the accident causation process using Structural Equation Modelling (SEM). They hypothesised that people's perceptions of various organizational processes and practices – what is now called "safety climate" - influence the state of safety in the organization and self report questionnaires can capture these perceptions. Studies of safety climate by Flin and her colleagues shed some light on the potential contributors to accidents, with climate measures capturing up to 50% of the variance in safety outcomes.

Other industries have replicated these findings and now accept that safety climate measures help to predict safety behaviours (Clark, 2006; Johnson, 2007). These findings appear to hold, whether the dependent variable is a self-reported measure of safety behaviour or actual measures of safety outcomes (e.g., Zohar, 2000; Cooper & Phillips, 2004; Johnson, 2007). Researchers in this area have therefore begun to pursue other lines of enquiry. One very active line of enquiry concerns the refinement of measurement instruments, which capture the essential elements of safety climate and there are now many well-validated instruments from which to choose (e.g., Seo, Torabi, Blair, & Ellis, 2004; Silva, Lima, & Baptista, 2004; Evans, Glendon, & Creed, 2007).

A second line of enquiry aims to establish the mechanisms by which climate influences safety behaviours. Working within this tradition, Fogarty (2004) employed a safety climate approach to assist in the development of a model to explain morale, psychological health, turnover intentions, and error in the aviation maintenance environment. An instrument called the Maintenance Environment Survey (MES) was constructed and administered to 240 personnel responsible for maintenance of a large military helicopter fleet. The structural model predicted 45% of the variance in psychological health, 67% of the variance in morale, 27% of the variance in turnover intentions, and 44% of the variance in self-reported maintenance errors. In a follow-up study, Fogarty (2005) administered a revised version of the MES to 150 aviation maintenance personnel to test the fit of a model in which the effect of safety climate on errors was partially mediated by individual level factors, such as psychological strain. He found support for the model and argued that in the efforts to secure better safety outcomes, a dual focus should be maintained on organizational and individual level variables.

Within this same tradition, other researchers have taken a broader approach. Neal and Griffin (2006) used a longitudinal design to explore the role of safety motivation as a potential mediator of the safety climate-safety behaviour relationship. They reaffirmed the connection between climate and behaviour but warned that it takes time for positive changes in safety climate and safety motivation to manifest themselves in lower accident rates. Among the recommendations flowing from their study was that researchers not treat safety behaviour as a unidimensional construct. They identified safety compliance and safety participation as examples of distinct constructs that are usually not separated in studies of safety outcomes. The present study adopts that same view, arguing that errors and violations are distinct safety outcomes that need to be treated differently.

The Current Study

Dekker (2003) noted that in modern usage error can mean three things: 1) error as the cause of failure (e.g., proficiency); 2) error as the failure itself (e.g., wrong decision); and 3) error as process, as an intentional departure from some kind of standard. Not distinguishing between these different possible definitions of error is a problem. To reduce this conceptual confusion, the authors propose to label this third category of errors as *violations*, a term already used by many researchers in this area. It is further proposed that errors and violations have different causes and that the distinction is therefore not simply a matter of nomenclature. In a broad sense, it has been said that errors tend to result from cognitive, social, and organizational factors, and violations tend to result from attitudinal, social, and organizational factors (e.g., Reason, 1995; Reason, 1997; Sutcliffe & Rugg, 1998). The proposition that errors and violations have different aetiologies is therefore not new, but often overlooked. Furthermore, the empirical evidence supporting this intuitive link is weak because most studies have focused on either errors (Fogarty, 2004, 2005) or violations (e.g., Lawton, Parker, Stradling, & Manstead, 1997; Mearns et al., 2001; Mearns, Whitaker & Flin, 2003). There is a need for studies that include both variables, linking them in a hypothetical nomological net that is testable using SEM techniques.

The measurement part of the model comprised four elements: indicators for Safety Climate, Psychological Health, Violations, and Errors. The structural part of the model comprised the hypothesised linkages between these four dimensions. Figure 1 shows the full model.
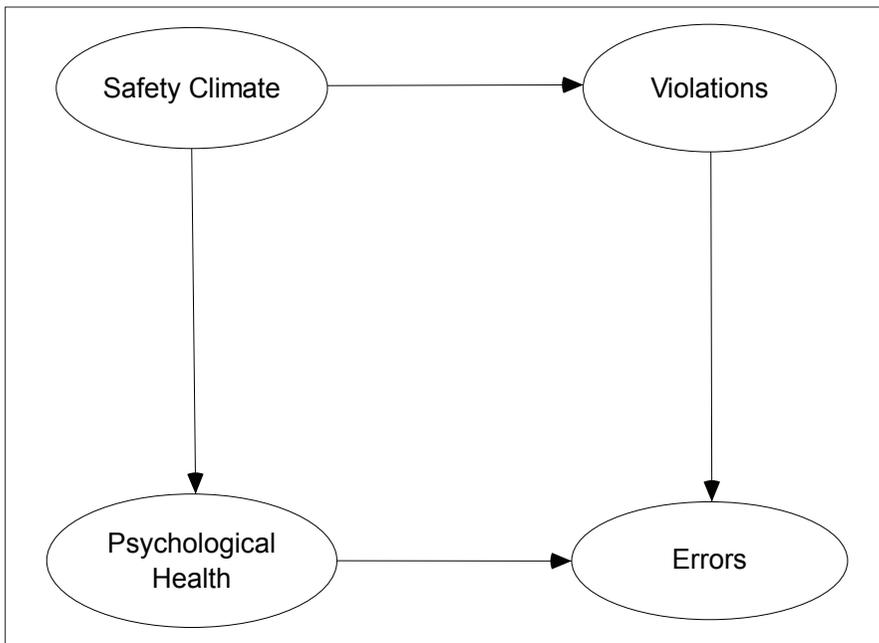

Figure 1. Conceptual model representing relations among Safety Climate, Psychological Health, Violations, and Errors (Indicator variables not shown)

In Figure 1, Safety Climate is represented by the reflective indicators Management Support, Commitment to Safety, Management's Awareness of Violations, Communication Effectiveness, Access to Resources, Training, and Workload. Fatigue, Strain and General Health Questionnaire (GHQ) are reflective indicators of the underlying construct called Psychological Health. Error Causes, Error Types, and Mistakes are reflective indicators of the construct Errors. Violation Attitude and Violation Behaviour are reflective indicators of a construct called Violations. Because most of the scales used in the present study were adapted from those already reported and validated in the literature, they were expected to define their hypothesised underlying dimensions.

The first part of the structural model comprises the direct link between Safety Climate and Psychological Health and a further direct link to Errors, thus modelling the indirect linkage between Safety Climate and Errors noted by Fogarty (2004, 2005). The second element in the structural model comprises the direct link between Safety Climate and Violations and a further direct link to Errors. In support of the first of these links, Helmreich (2000) suggested that violations can stem from a culture of non-compliance, perceptions of invulnerability, or poor procedures. He also reported that over half the "errors" observed in a line audit safety operations (LOSA) exercise were due to violations and that those who violated procedures were more likely to commit other types of errors. Mearns et al. (2001) found pressure for production and work pressure explained 58% of the variance in a construct they labelled Safety Behaviours, with pressure for production being the main contributor. Scales measuring violations often appear in the literature as safety behavior scales, so this finding supports the direct link between Safety Climate and Violations. Other researchers have confirmed this link (e.g., Neal, Griffin, & Hart, 2000; Oliver, Cheyne, Tomás, & Cox, 2002; Rundmo, 2000; Rundmo, Hestad, & Ulleberg, 1998). The final link, between Violations and Errors, is strongly supported by the literature, which shows procedural violations are the best predictors of accident involvement (Hofmann & Stetzer, 1996; Lawton, Parker, Stradling, & Manstead, 1997; Lawton & Parker, 1998; Meadows, Stradling, & Lawson, 1998; Mearns et al., 2001).

We labelled this Model 1, the fully mediated model. Two competing models were also tested. Reason (1997) proposed that workplace conditions (safety climate factors) cause unsafe acts :such as inadequate tools and equipment, undue time pressure, insufficient training, under-staffing, poor supervisor-worker ratios, and unworkable procedures. Therefore, Model 2 differed from Model 1, it included an additional pathway from Safety Climate to Errors. We called this the partially mediated mode. Model 3 was also a minor variation of Model 1 with a pathway fitted between Psychological Health and Violations to test whether the previously-noted direct effects of health on errors (Fogarty, 2004, 2005) extends to other forms of safety behaviours.

Method

Participants
Three hundred eight maintenance personnel from the Australian Defence Force (ADF) were involved in the study. Of the personnel who completed the survey, 33.7% (N = 105) were from the Army, 27.6% (N = 86) from the Navy, and 37.2% (N = 116) from the Air Force.

The Survey Instrument

Subject matter experts from the Australian Defence Force (ADF) participated in the development of the Flying Safety in Maintenance Climate Survey for the present study. The survey was divided into eight sections: (a) Background Information, (b) Flying Safety, (c) Workplace Flying Safety, (d) Working Procedures and Practices, (e) Reporting Procedures and Practices, (f) Training and Resources, (g) Other Issues and (h) General Health. Some sections of the survey were of interest to the Directorate of Flying Safety but not to the authors. The subscales described below are those relevant to the current study.

There were seven subscales in the Safety Climate section of the survey.
1. Management support (Mgntsup), where three items measured how often management listened to safety concerns from subordinates such as supervisors and tradesmen (e.g., Managers listen to concerns from tradesmen/supervisors and react appropriately);
2. Safety commitment (Safecomm), where four items assessed how committed the organization, management, and colleagues were to safety (e.g., The ADF is committed to flying safety);
3. Management's awareness of conditions affecting safety (Mgntawar), where three items assessed management's awareness of workplace pressure and resulting shortcuts (e.g., Managers are aware that the pressure placed on supervisors makes it necessary to take shortcuts/risks to achieve the task);
4. Communication effectiveness (Commeff), where three items measured the extent to which management was successful in communicating safety issues to subordinates (e.g., Management communicates issues effectively to tradespersons);
5. Access to resources (Resacc) where four items assessed the availability of various resources such as personal protective equipment, manuals, equipment, and tools (e.g., I have access to all the tools that I need for my work);
6. Training standards (Train) where seven items were used to assess the adequacy of training, including on-the-job training, trade skills, systems knowledge and formal training (e.g., The trade skills of junior personnel are adequate);
7. Workload (Workload), which was assessed using five items that rated the complexity of task performance (e.g., I undertake tasks concurrently to get the job done).

The Safety Climate items mostly employed 5-point ratings that ranged from 1 (strongly disagree) to 5 (strongly agree). Resources, Workload, and Management Support were rated on a 5-point scale that ranged from 1 (always) to 5 (never). Scores were recoded so that higher scores indicating a higher level of resources, workload and management support.

After the climate section, three subscales measured the latent construct Psychological Health. The first of these was an abbreviated version of the strain scale used in Fogarty (2004, 2005). It comprised five items (e.g., How often do you feel stressed at work because of the job itself?). Four items were included to measure

fatigue (e.g., How often do you feel fatigued at work because of the working hours?). Response options for both the strain and fatigue subscales ranged from 1 (never) to 5 (always). The third subscale was the 12-item version of the General Health Questionnaire (GHQ: Goldberg and Williams, 1988). The GHQ explores several aspects of psychological health and measures job-related strain (Parkes, 1992; Payne, Wall, Borrill, & Carter, 1999). Participants were required to respond to a number of statements regarding the state of their psychological health: anxiety and insomnia (e.g., Lost much sleep over worry?); social dysfunction (e.g., Have you felt that you are playing a useful part in things?); and severe depression (e.g., Been thinking of yourself as a worthless person?). Scores on this variable were recoded so that higher scores indicated better psychological health.

The next section of the survey instrument used two subscales to measure procedural violations. In the first of these (Violbeh), comprising five items, respondents indicated how frequently they engaged in unsafe behaviours (e.g., I will temporarily disconnect or remove apart to make a job easier, but not document the disconnection/removal). Possible responses ranged from 1 (always) to 5 (never). The second subscale (Violatt), comprising four items, tapped willingness to violate rules and procedures (e.g., I am prepared to take risks, other than those inherent in my job, to get a task done). Violatt employed a 5-point Likert scale that ranged from 1 (strongly disagree) to 5 (strongly agree). For both of these subscales, scores were recoded so that higher scores denoted a higher occurrence of violations or a greater willingness to engage in procedural violations.

In the final section of the survey, items from the Maintenance Environment Scale (MES: Fogarty, 2004) and the 48-item aircraft maintenance checklist developed by Hobbs and Williamson (2000) were used to form three marker variables for the latent construct, Errors. The first subscale (Errtype, 10 items), asked respondents to indicate how often they made different types of errors (e.g., I have missed out steps in maintenance tasks). In the second subscale (Errcaus, 10 items), respondents were required to indicate how often they had made errors because of different background factors (e.g., I make errors because of lack of concentration). In the third subscale (Mistakes, 4 items), respondents indicated how often they made mistakes due to training deficiencies (e.g., I make mistakes because my systems knowledge is lacking). Ratings for all subscales were made on a 5-point scale ranging from 1 (always) to 5 (never). Scores were recoded so that higher scores represented the occurrence of more errors and mistakes.

For all subscales, the dependent variable was the mean response for the subscale, that is, the total score divided by the number of items.

Procedure
Serving members of the ADF Directorate of flying safety administered the survey to participant groups. Maintenance workers, maintenance officers, and personnel indirectly related to maintenance work were asked to participate in this study. The surveys were completed in group sessions lasting from 30 to 45 minutes and were then mailed to the university research team.

Statistical Analyses

The competing structural equation models were proposed and tested using the AMOS 4.0 (Arbuckle, 1999) program. Model fit was assessed using Chi Square (χ^2), the Chi Square to degrees of freedom ratio (χ^2/df), the Tucker-Lewis Index (TLI), the comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA).

Results

A small number of missing values were replaced using the expectation-maximisation (EM) algorithm (Roth, 1994) in SPSS, statistical software, version 10.0. Following data screening, descriptive statistics were compiled to ascertain the spread of scores on the indicator variables. The means and standard deviations show a reasonable spread of scores. Additional normality checks (not reported) showed positive skewness on safety commitment (Safecomm), access to resources (Resacc), and the two measures of violation behaviours. GHQ scores were negatively skewed. These outcomes were not surprising, and the degree of skewness was not judged problematic for the multivariate analyses to follow. With the exception of the training subscale, the internal consistency reliability estimates (Cronbach's alpha) for all variables were above .70, and most were above .80.

The main aim of the study was to test the conceptual model shown in Figure 1 and to compare fit indices with those obtained for two competing models. These fit indices for these three models are summarised in Table 1.

Table 1
Summary of Fit Statistics for Different Models

Model	χ^2	df	p	χ^2/df	TLI	CFI	RMSEA
1	182.66	86	< .01	2.12	.92	.94	.06
2	182.62	85	< .01	2.15	.92	.94	.06
3	180.89	85	< .01	2.13	.92	.94	.06

The fit indices for all three models were indistinguishable in terms of their fit to the data and were either on the borderline or within commonly recommended cut-off values for these fit indices. Model 1, the fully mediated model, gave a slightly more parsimonious account of the data, however, so we selected it as our preferred model. Figure 2 shows the full measurement and structural model, with parameter estimates.
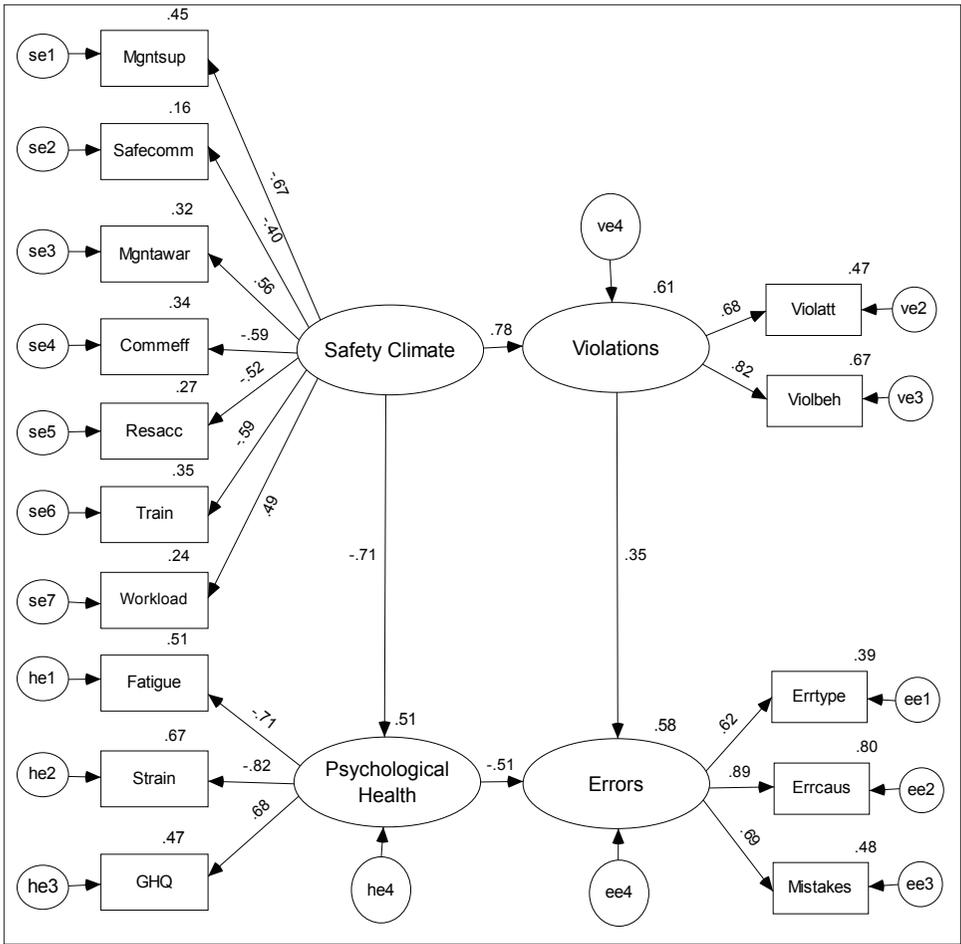
Figure 2. Empirical model representing relations among Safety Climate, Psychological Health, Violations, and Errors

All pathways shown in the model were significant. The model accounted for 51% of the variance in Psychological Health, 61% of the variance in Violations, and 58% of the variance in Errors. As well as the direct effects, there was a significant indirect effect of Safety Climate on Errors (*b* = .65, p < .01).

Discussion

The aim of this study was to validate and extend existing models of organizational and individual factors in the prediction of unsafe acts. The study brought together the key outcome variables of errors and violations and related these to organizational and individual factors in a model, which described the direct and indirect effects of safety climate and individual psychological health on self-reported errors and violations. The outcomes support the claims of other researchers that safety climate directly influences violations (e.g., Oliver et al., 2002; Rundmo, 2000; Rundmo et al., 1998), and that individual health directly influences the frequency of errors (Fogarty, 2004, 2005). Specifically, a large amount of the variance

in violations (63%) can be explained by the safety climate of the organization and a large amount of the variance in errors (58%) can be explained by the combined effects of safety climate and psychological health. This study has supported the proposition that errors and violations have different psychological antecedents.

These findings are important to safety practitioners, particularly in the aviation industry. Hudson (2007) has written a very useful road map for implementing a safety culture in an organization. Towards the end of the paper, he warns academics against the dangers of continuing to refine measurement instruments instead of looking at how the instruments are used and what he calls coming "down from the trees" (p. 719) and engaging with industry. At the same time, he emphasises the importance to industry of having well-founded empirically justified theories. We would like to think that our focus in this study on breaking down safety problem behaviours into two easily-recognised components and showing that they have different aetiologies places us near to the bottom of the tree. A one-size-fits-all approach to safety behaviours might well prove effective but it will be highly inefficient. Attempts to reduce intentional and unintentional unsafe acts should be aimed at both individual and organizational levels, with an understanding of the different origins of errors and violations.

Whilst these findings replicate earlier research on errors and break new ground by considering errors and violations together, we should point out that the methodological shortcomings in this study. Firstly, using cross-sectional methodology is an evidently weak approach to causality (MacCallum & Austin, 2000). The use of self-report measures for all variables is also problematic in that there is the possibility of method variance as the source of commonality among the variables. One global concern of studies that involve structural equation modelling is that conclusions are likely to be limited to the particular sample. In this study, a restricted sample was used, that is, military aviation maintenance, and results should be treated cautiously when generalising beyond this population as the military population may not be representative of the maintenance population in general.

Safety climate measures such as the ones used in these studies are very useful but they should be standardised so that the items and scales are the same across administrations, thus permitting the establishment of benchmarks on the various scales (Flin, Mearns, O'Connor, & Bryden, 2000). The focus of the research up to this point has been the identification of key safety outcomes and defining the network of relationships among these variables and background climate variables. A further aim has to be the linking of self-report measures with actual performance outcomes, rather than simply using self-report as the basis of measurement operations. The low base rates of incidents and errors suggests that this research will involve higher level modeling, but it is our expectation that the models developed to this point will prove useful in explaining safety data, whatever form it takes.

In conclusion, the safety literature tends to be dominated by discussions of error taxonomies and descriptive models of accident causation, such as the Reason (1997) model. We see these contributions as valuable but we also believe that they must be supported by empirical research. Structural equation modelling is a technique that can be used to test assumptions embedded in popular descriptions of accident causation. This study has developed and validated a model that encompasses a number of organizational, social and individual factors that predict a significant proportion of the variance in self-reported errors and violations. Ongoing studies, seek to extend the model presented here to include incident reporting, another key psychological variable in the quest to achieve safer and more productive working environments. Safety will continue to be critical as complex high-risk industries, such as aviation, become more technologically driven and complicated. Consequently, organizations will need to maintain a heightened awareness of safety, risk, and security.

Acknowledgements

References

Arbuckle, J. L. (1999). Amos users' guide. Version 4.0. Chicago, IL: SmallWaters Corporation.

Clark, S. (2006). The relationship between safety climate and safety performance: A meta-analytic review. Journal of Occupational Health Psychology, 11(4), 315-327.

Cooper, M.D., & Phillips, R. A. (2004). Exploratory analysis of the safety climate and safety behaviour relationship. *Journal of Safety Research, 35*, 497-512.

Dekker, S. W. (2003). Illusions of explanation: A critical essay on error classification. *International Journal of Aviation Psychology, 13*(2), 95-106.

Evans, B., Glendon, A. I., & Creed, P. A. (2007); Development and initial validation of an aviation safety climate scale. *Journal of Safety Research, 38*, 675-682.

Fleming, M., Flin, R., Mearns, K., & Gordon, R. (1998). Risk perceptions of offshore workers on UK oil and gas platforms. *Risk Analysis, 18*(1), 103-110.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 3*(1-3), 177-192.

Fogarty, G. (2004). The role of organizational and individual differences variables in aircraft maintenance performance. *International Journal of Applied Aviation Studies, 4*(3), 73-90.

Fogarty, G. (2005). Psychological strain mediates the impact of safety climate on maintenance errors. *International Journal of Applied Aviation Studies, 5*(1)*,* 53-63.

Goldberg, D., & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. Windsor, Berkshire: NFER-Nelson.

Helmreich, R. L. (2000). On error management: Lessons from aviation. *British Medical Journal, 320*, 781-785.

Hobbs, A., & Williamson, A. (2000). Aircraft Maintenance Safety Survey – Results. Canberra: Department of Transport and Regional Services.

Hofmann, D., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviours and accidents. *Personnel Psychology, 49*(2), 307-339.

Hudson, P. (2007). Implementing a safety culture in a major multi-national. *Safety Science, 45*, 697-722.

James, W. (1890). *The principles of psychology*. New York: Holt.

Johnson, S. E. (2007). The predictive validity of safety climate. *Journal of Safety Research, 38*, 511-521.

Lawton, R., & Parker, D. (1998). Individual Differences in Accident Liability: A review and integrative approach. *Human Factors, 40* (4), 655-671.

Lawton, R., Parker, D., Stradling, S., & Manstead, A. (1997). Predicting road traffic accidents: The role of social deviance and violations. *British Journal of Psychology, 88*, 249-262.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation Behaviour in psychological research. *Annual Review of Psychology, 51*, 201-226.

Meadows, M. L., Stradling, S., & Lawson, S. (1998). The role of social deviance and violations in predicting road traffic accidents in a sample of young offenders. *British Journal of Psychology, 89*, 417-431.

Mearns, K., & Flin, R. (1999). Assessing the state of organizational safety – Culture or climate? *Current Psychology: Developmental, Learning, Personality, Social, 18*(1), 5-17.

Mearns, K., Flin, R., Gordon, R., & Fleming, M. (2001). Human and organizational factors in offshore safety. *Work and Stress, 15*(2), 144-160.

Mearns, K., Whitaker, S. M., & Flin, R. (2003). Safety climate, safety management practice and safety performance in offshore environments. *Safety Science, 41*, 641-680.

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual Behaviour. *Safety Science, 34*, 99-109.

Neal, A., & Griffin, M.A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behaviour, and accidents at the individual and group levels. *Journal of Applied Psychology, 91*(4), 946-953.

Oliver, A., Cheyne, A., Tomás, J. M., & Cox, S. (2002). The effects of organizational and individual factors on occupational accidents. *Journal of Occupational and Organizational Psychology, 75*, 473-488.

Parkes, K. R. (1992). Mental health in the oil industry: A comparative study of onshore and offshore employees. *Psychological Medicine, 22*, 997-1009.

Payne, R. L., Wall, T. D., Borrill, C., & Carter, A. (1999). Strain as a moderator of the relationship between work characteristics and work attitudes. *Journal of Occupational Health Psychology, 4*, 3-14.

Reason, J. (1995). A systems approach to organizational error. *Ergonomics, 38*(8), 1708-1721.

Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Aldershot, Hampshire, England: Ashgate.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*, 537-560.

Rundmo, T., Hestad, H., & Ulleberg, P. (1998). Organisational factors, safety attitudes and workload among offshore oil personnel. *Safety Science, 29*, 75-87.

Rundmo, T. (2000). Safety climate, attitudes and risk perception in Norsk Hydro. *Safety Science, 34*, 47-59.

Seo, D-C., Torabi, M.R., Blair, E.H., & Ellis, N.T. (2004). A cross-validation of safety climate scale using confirmatory factor analytic approach. *Journal of Safety Research, 35*, 427-445.

Silva, S., Lima, M.L., Baptista, C. (2004). OSCI: an organisational and safety climate inventory. *Safety Science, 42*, 205-220.

Sutcliffe, A., & Rugg, G. (1998). A taxonomy of error types for failure analysis and risk assessment. International *Journal of Human-Computer Interaction, 10* (4), 381-405.

Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology, 85*, 587–596..

Investigation into Accident Initiation Events by Flight Phase, for Highly Inexperienced Glider Pilots.

Steve Jarvis and Don Harris

*Department of Systems Engineering and Human Factors,
School of Engineering
Cranfield University
Cranfield, Bedford MK43 0AL
United Kingdom*

Abstract

Despite inexperienced glider pilots having accidents at a far higher rate than average, it is unknown if accidents sustained by inexperienced pilots stem from different elements of the flight from those sustained by all other more experienced pilots. All pilot-related accidents in the British Gliding Association database (2002 – 2006) were identified and accident rates were calculated for each flight phase using exposure data derived from previous research, thereby allowing comparisons to be made on a per-flight basis. The results showed that for inexperienced pilots (10 hours and under), accidents associated with a seminal event in the approach phase occurred at 8 times the rate than for experienced pilots. For seminal events in the landing phase accidents occurred at 6 times the rate. This research demonstrates the benefit of examining the flight phase in which the seminal event of an accident occurs for better targeting of appropriate remedial actions (rather than classifying the accident by flight phase) and of the use of flying exposure measures to produce comparable accident rate data rather than simply comparing accident counts.

Investigation into Accident Initiation Events
by Flight Phase, for Highly Inexperienced Glider Pilots.

The estimated proportion of accidents caused by pilot factors remains high for all types of aviation. Feggetter (1982) reported that the figure was approximately 70%. Studies that are more recent have reported figures of the same magnitude, for example 71% (O'Hare, 1994); 70% (The Bureau of Air Safety Investigation – BASI, 1996); and 78.6% (Aircraft Owners and Pilots Association, 2006). Research into US gliding accidents reported that 79.7% were pilot related (van Doorn & de Voogt, 2007).

Research on pilot experience in relation to accident involvement has shown conflicting evidence. Experience brackets (for example 200 – 400 hours) have

been reported as equating to a higher chance of accident involvement (Booze, 1977, Jenson, 1995, Olsen & Rasmussen, 1989). This implies that highly inexperienced solo pilots are not the most likely to have an accident. Other studies have dismissed such claims as being unfounded, or based on unsound statistics (Li, 1994, O'Hare & Chalmers, 1999). A recent study of general aviation accidents by the Air Safety Foundation of the Aircraft Pilots and Owners Association (2006) concluded that pilots with fewer than 200 hours total time are the most vulnerable and those with fewer than 10 hours in make and model are more vulnerable still. In gliding, a simple frequency distribution of all UK accidents from 1997 – 2006 strongly suggests that glider pilots with 10 or fewer hours as pilot in command (PIC) are a particularly vulnerable group (Figure 1). A study using flight exposure estimates showed that very inexperienced solo glider pilots (those with fewer than 10 hours experience as PIC) have twice the number of accidents per launch and three times the number of accidents per hour flown than their more experienced counterparts (Jarvis & Harris 2007).
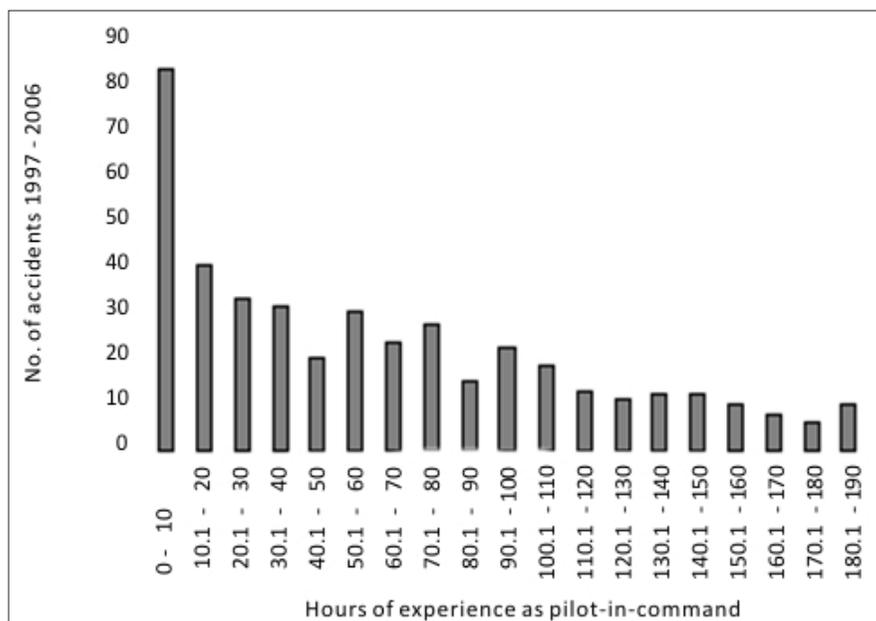

Figure 1. Frequency distribution of all UK gliding accidents 1997 – 2006 (data provided by British Gliding Association).

The reasons for these differences in accident rate are not known; however, an analysis of the flight phases in which accidents occur may provide some insight. A study of 143 US gliding accidents by van Doorn & de Voogt (2007) found that based simply on a frequency analysis, over half of all accidents (52.4%) occurred on landing, while 30% occurred during the cruise. Most fatal accidents in gliders take place during the cruise (van Doorn & Zijlstra, 2006); 36% of these accidents were found to end in fatality, whereas only 10% of landing accidents did so (van Doorn & de Voogt, 2007). These studies did not, however, break down the accident statistics by pilot experience level nor did they take into account flight exposure.

In general aviation, landing and maneuvering are consistently cited as being the most common accident flight phases. Again, based solely on frequency counts, O'Hare (1994) found that 39% of accidents occurred during the landing phase. AOPA (2006) reported that 38.9% of non-fatal accidents happened during landing, more than any other flight phase, while observing that maneuvering was the most common flight phase for fatal accidents (22.8%). In a study of insurance claims, Lenné & Ashby (2006) reported that the landing and taxiing phases accounted for 55% of all non-fatal general aviation accidents in Australia.

The classification of accidents by any single flight phase is far from straightforward. Aviation accidents are often the result of a chain of events rather than a single event or error (Wiegmann et al, 2005). Therefore, taxonomies and classifications that categorize accidents by a single flight phase may risk over-simplification, as a series of causal events may have accumulated during the flight. This scenario is particularly likely in gliding because of the difficulty in regaining lost energy in terms of height and/or speed. For example: misjudgment of height when entering the circuit may lead to poor positioning of the base leg with little energy to reach the airfield, subsequently resulting in a slow approach and heavy landing. In such a case the accident *flight phase* might be categorized as the stage in which the damage/injury was sustained (landing) thereby failing to identify that the initiating event occurred much earlier in the flight. Recognizing the issue of multiple events Wiegmann et al. (2005) categorized accidents using any number of flight phases but labeled only one of these to be the *seminal* phase, in which the initiating event was deemed to have taken place. This same approach was initially used in the study of North Sea Helicopter Safety (Ingstad et al, 1990). Lenne & Ashby (2006) also used a similar method by identifying the first crash occurrence noted in the accident narrative. Furthermore, some flight phase analyses can also be criticized as lacking explanatory power because of the nature and extent of the phases used. Van Doorn & de Voogt (2007) used four phases to describe all accidents: *assembly, tow, cruise,* and *landing*. Although problems can occur during assembly, it is problematic to compare this numerically with other phases of flight since many gliders are kept in hangers or flown many times per assembly; hence, many glider flights do not include this phase at all. This leaves only three in-flight phases, all of which include numerous flight components (see British Gliding Association, 2003; Stewart 1994; Piggott, 1997) meaning that categorization using such taxonomy is questionable in terms of its utility in identifying specific problem areas. For example, an accident deemed to have occurred in the cruise phase could have taken place in midair (e.g. a collision or overstress leading to break-up), in the circuit or approach to a field landing, during an attempted out-field landing, or as a result of unintentional ground impact. An analysis with more explanatory power allowing better-targeted remedial interventions is required, not just simply using a more detailed breakdown of flight phases but also paying more attention to causal (seminal) events rather than to the final event.

Although research dedicated to accident flight phase has been conducted for both general aviation and gliding, there has been little attention to the relationship between pilot experience and the flight phases of initiating accident events. Fur-

thermore while some research has used frequency data and accident totals (van Doorn & Zijlstra, 2006; van Doorn & de Voogt, 2007) no research has provided comparable accidents rates, due to the difficulties of obtaining exposure data for pilots of differing experience levels. Using UK gliding accident data, this study compares the flight phases in which the initiating event preceding an accident occurred with respect to highly inexperienced pilots (10 or fewer hours experience as PIC) and more experienced pilots (over 10 hours). Furthermore, the data obtained are used to provide estimates of accident rates (both in terms of hours flown and number of launches) for both phase of flight and pilot experience.

Method

The analysis of accident data progressed in three stages. Firstly, all UK gliding accidents from 2002 - 2006 deemed to have pilot-related causes were identified. Secondly, the number of pilot-related events were identified within each accident report with the *seminal event* being categorized as the first to occur. Lastly, the seminal events were subjected to a flight phase analysis, using a detailed template. All accidents were analyzed according to two levels of pilot experience (10 or fewer hours experience as PIC and more than 10 hours PIC), in line with the research findings by the Air Safety Foundation of the Aircraft Pilots and Owners Association (2006) and Jarvis & Harris (2007), showing that 10 hours PIC or fewer is a particularly vulnerable experience bracket. Since pilot experience level was a key variable of concern in the research, this information was removed from the accident descriptions during categorization to avoid influencing the process.

Data
The British Gliding Association (BGA) database provided the data of all UK gliding accidents and incidents over five years from 2002 to 2006 (source: British Gliding Association, 2007). This database contained details including pilot age (years); total experience in command (hours); aircraft type; severity of injuries (none, minor, serious, fatal); damage to the glider (none, minor, substantial and write-off); and a narrative description of what happened. As an initial step, accidents resulting in no injury or damage were dropped from the analysis. Additionally, ground-handling accidents (such as towing out winch cables, or pulling gliders out of hangers) were also omitted from the analysis. Where possible the accident descriptions contained within the BGA database were supplemented with segments from AAIB (Air Accident Investigation Branch) or longer BGA accident reports.

Stage 1: Identification of pilot-related accidents. All accidents from the BGA database (2002 – 2006) were categorized into either primarily *Pilot-Related* or *Other* cause (*Technical, External or Unknown*) using a set of guidelines drawn up and agreed to by a group of subject matter experts and an aviation human factors professional. All members of this group were experienced instructors on gliders and general aviation aircraft, with a combined experience of over 10,000 hours of logged flying time.

The definition of a Pilot-Related cause was based upon Hollnagels' definition of human error (Hollnagel, 1998). To be defined as a Pilot-Related cause there had to be an identifiable performance shortfall in terms of the actions (or inactions) on

the part of the PIC together with a reasonable opportunity for the pilot to act in such a way that could have avoided the accident. If a Technical or External factor was identified as being the seminal accident event then the accident was considered as non-pilot related (i.e. placed in the Other category).

The guidelines for coding an accident as being the result of a *Technical Factor* were that the aircraft would have been deemed unserviceable had the failure been apparent before flight. If a Technical Factor was induced by the abnormal operation of the glider (outside its operating limitations) this was deemed to be pilot induced. An *External Factor* was regarded as any reasonably unforeseeable and/or unavoidable factor external to the glider that made the flight difficult beyond the skills that could reasonably be expected of a competent pilot. External Factors brought about by pilot actions, inactions or decisions (that were reasonably foreseeable) were again deemed to be Pilot-Related. Furthermore, difficult flying conditions were only counted as External Factors where there were no reasonable signs or expectation of such conditions occurring. A lack of rising air (thermal, wave, or ridge lift) was not regarded as an External Factor since such lift is not reliable and it is also not required for safe glider operation. If it was not possible to identify positively any Pilot-Related, Technical or External Factor (i.e. where no causal events could be determined by the rater) the accident cause was categorized as being *Unknown*.

Stage 2: Identification of number of pilot-related contributory events. The number of major *pilot contributory* events in each accident was identified from the accident narrative. Following this, the seminal event was identified; this being defined as the first event in the sequence (cf. Ingstad, et al., 1990; Wiegmann et al, 2005).

Stage 3: Flight phase analysis. A high-level mission analysis utilizing concepts drawn from process charting methods (see Kirwan & Ainsworth, 1992) was undertaken to breakdown the operation of a glider into meaningful, quasi-independent flight phases. Resources such as the BGA instructor's manual Edition 2 (2003), Piggott (1997) and Stewart (1994) were used in this process along with a number of subject matter experts (experienced gliding instructors). This analysis was performed to produce a two-level flight phase template. The resulting template consisted of 25 flight phases in total, grouped within six higher-order phases (pre-flight; launch; in-flight phase; circuit; approach and landing). Agreement was reached between the subject matter experts that the final template was representative of all aspects of UK gliding operations. This coding template is shown in Figure 2.

```
1.    Pre-flight, after boarding

2.    Launch
      a.    Ground run [Only if attached to cable]
      b.    Pre-rotation initial climb (Winch)/ Airborne, pre-climb (Aerotow)/PIOs
      c.    Rotation and establishing main climb (Winch)/Initial Climb (Aerotow)
      d.    Established Climb
      e.    Aerotow cruise (cross-country towing etc)
      f.    Release
      g.    Recovery to normal flying speed, period prior to manoeuvre/approach
      h.    Non normal aerotow: Low tow, 'Boxing the tug'

3.    Flight phase
      a.    General flying (practicing manoeuvring, local soaring/flying etc)
      b.    Serious soaring (circling, dolphin, eights, street flying, cloud climb, leaving,
            joining etc)
      c.    Search/descent (Usually during cross-country/ extended soaring. Search for
            lift / search and inspection of field to land, prior to circuit/ abbreviated circuit
            commitment)
      d.    Ridge soaring
      e.    Wave flying
      f.    Final glide (incl comp finish/ return to airfield/ 'stretching the glide')
      g.    Aerobatics/ intentional unusual manoeuvring (stall/spin/steep turns)
      h.    Immediate evasive airborne manoeuvre to avoid imminent collision

4.    Circuit
      a.    Circuit join. [Phase between 3 & circuit to land] (including prep, wheel down,
            decisions on circuit/landing direction). NOT choice of field itself, that is 3c
            (search/cruise)
      b.    Circuit (from high key to final turn, joined from anywhere) Include as seminal
            where accident descriptions begin from a poorly positioned
            final turn or final approach from turn  (too high, low, far, close)
      c.    Abbreviated circuit or non-standard manoeuvring to land
      d.    Final turn (from normal or abbreviated circuit only)

5.    Approach
      a.    Approach after circuit or flying (approach other than 4b)
      b.    Approach as straight ahead recovery from launch failure

6.    Landing/Ground
      a.    Flare/ Hold off. Includes ballooning and PIOs
      b.    Ground run [post landing or after ground cable release]
```

Figure 2. Flight phase categorization template

     The flight phase analysis highlighted the requirement to separate accidents occurring during attempts to land at an airfield from accidents occurring while attempting to land in an unfamiliar field, which can often occur when insufficient lift is found to continue the flight. Off-airfield landings (also known as field landings) are common in gliding but involve unique tasks such as assessing field size and suitability, and positioning a circuit to an unfamiliar site with no primary height information. This is accepted as common practice in gliding (rather than an emergency) particularly when a pilot is attempting a cross-country soaring flight. It was therefore necessary to be able to identify such accidents during analysis in case they had a substantial effect on the findings. Therefore, prior to attaching the flight phase descriptors, each accident was classified by its location (*airfield* or *off airfield*). Accidents in the circuit or approach phase of the base airfield (or intended landing airfield) were labeled as airfield accidents, whereas those occurring outside the circuit pattern of the airfield were treated as off-airfield accidents.

     Accidents following launch failures required identification for similar reasons. Launch failures can require unique maneuvers such as regaining flying speed at

low altitude and flying low abbreviated circuits. Therefore, it was essential to have the ability to separate these from normally launched glider flights.

All accident seminal events were categorized using the flight phase template (figure 2). In addition, each accident was further categorized as being normal launch/launch failure; airfield/off-airfield.

Reliability of the ratings
In accordance with previous researchers using large samples of accident data (e.g. Gaur, 2005), to establish inter-rater reliability a random sample of 100 accidents was independently categorized by the primary investigator and an independent rater. The latter was an experienced pilot of gliders and general aviation aircraft and was an airline training captain and crew resource management (CRM) instructor with training in human factors. In order to check observer consistency, the sub-sample was re-categorized by the primary investigator two weeks later to establish the intra-rater reliability (a factor omitted in many studies).

Cohen's Kappa was used to calculate inter- and intra-rater reliability at each stage of the analysis. Robson (2002) suggests Kappa values ranging from 0.6 – 0.75 are *good*, and above 0.75 are *excellent*. The results for both measures of rater reliability were *good* or *excellent* at all stages: Identification of pilot-related accidents; $Kappa_{(inter-rater\ reliability)}$ = 0.87, $Kappa_{(intra-rater\ reliability)}$ = 0.88: The number of pilot-related contributory events; $Kappa_{(inter-rater\ reliability)}$ = 0.61, $Kappa_{(intra-rater\ reliability)}$ = 0.61: Flight phase analysis; $Kappa_{(inter-rater\ reliability)}$ = 0.79, $Kappa_{(intra-rater\ reliability)}$ = 0.90. This indicated that categorization of accidents was both reliable and consistent at all stages of analysis.

Results

Data analysis
Initially, Fisher's exact tests were used to establish if significant differences existed between the inexperienced and experienced groups in terms of the simple proportions of pilot related accidents, as well as other factors such as injury severity and aircraft damage. The distribution of seminal events occurring during the various flight phases with respect to pilot experience groups were also analyzed in this way. Odds ratios with associated confidence intervals were calculated between the two experience groups for all six top-level flight phases.

Although such analyses can be used to compare the frequency of accidents in one group with that of the other and identify where accident features were disproportionately distributed between groups, they cannot account for differences in flying exposure between the groups (including flying that did not result in an accident). For this, accident rates were required.

The exposure estimate from Jarvis & Harris (2007) was recalculated using the same method but including two additional years of data in order to cover the period 2002 to 2006 inclusive. The 10 or fewer hours exposure estimate was

subtracted from BGA annual totals to provide data for the two groups; pilots with 10 or fewer hours and those with more than 10 hours as PIC. On this basis the estimated total number of launches taken from 2002 – 2006 by pilots with 10 or fewer hours PIC was 29,924 with an upper 95% confidence boundary of 35,301 launches and a lower 95% confidence boundary of 24,548 launches. The estimated total number of hours flown was 11,553 hours (upper 95% confidence boundary of 14,017 hours and a lower 95% confidence boundary of 9,089 hours). The mean calculation of flying exposure by pilots with over 10 hours PIC during the same period was 1,609,810 launches and 696,041 hours.

Stage 1: Identification of pilot-related accidents, and pilot related events within each accident. Of 469 accidents, no causal factors could be determined for 19, hence these were eliminated from the analysis. Of the remaining 450 accidents, 418 occurred to pilots with over 10 hours PIC, of which 331 were deemed to have been pilot-related. For pilots with 10 or fewer hours PIC there were 32 accidents, 28 of which were designated pilot related. A Fisher's exact test on this data was non-significant (p = 0.361) suggesting that the distribution of causes (Pilot-Related or Technical or External) was randomly distributed among pilots across the two levels of experience.

Table 1 shows the number of accidents leading to injury and damage for the two pilot experience groups. Fishers exact tests on these data show no significant association between the degree of injury and experience group over the five years being studied (p = 0.701). The same is true of aircraft damage analyzed by experience group (p = 0.272). There is therefore no evidence to suggest that the accidents suffered by inexperienced pilots were different in terms of their consequences to those suffered by more experienced pilots.

Stage 2: Identification of number of pilot-related contributory events. Within the 359 pilot-related accidents, 545 causal events were identified in total. Only three accidents were categorized as containing four events, hence these were combined with the three-event accident group (see table 1). The resulting analysis using a Fisher's exact test showed a significant association between pilot experience group and the number of events in the accident sequence (p = 0.016). Further analysis of standardized residuals indicated that the 10 or fewer hours group had a significantly higher proportion of accidents where three or more events were identified in the analysis (standardized residual of 2.3, p = 0.016).

Table 1
Total accident numbers for the two experience groups broken down by injury severity, damage classification, the number of identified events in the report narrative, and the location of the accident

		Under 10	Over 10	Total
Injury	Fatal	0	9	9
	Serious	0	18	18
	Minor	3	40	43
	None	25	264	289

Damage	Write off	0	36	36
	Substantial	11	107	118
	Minor	17	186	203
	None	0	2	2
No. of events	1 event	11	203	214
	2 events	10	97	107
	3 or more events	7	31	38
Accdent location	Airfield / within circuit	27	219	246
	Off-airfield	1	112	113

Stage 3: Flight phase analysis. Twenty-four accidents were caused by launch failures. Of these 23 occurred to pilots with over 10 hours of experience and one to a pilot in the 10 or fewer hours group. A Fisher's exact test on this data gave a two-tailed result of p = 0.707, showing that a systematic effect was improbable. No further analyses were done on these data.

Table 1 shows that of the 113 accidents identified as occurring off the airfield (meaning outside the circuit) all but one involved pilots with over 10 hours flying experience. The result of the Fisher's exact analysis suggested that accidents occurring away from the base airfield were not randomly distributed across the experienced and inexperienced groups (p < 0.000, two tailed). In terms of simple frequency of occurrence, experienced pilots were much more likely to have an accident away from their home airfield. The odds ratio was 13.807, suggesting far greater odds of experienced pilots sustaining this type of accident, although the 95% confidence interval is extremely wide (1.85 – 102.9).

The distribution of seminal event occurrence across the six high-level flight phases (broken down by pilot experience) is shown in table 2. It shows that of the six high-level phases general flying included most seminal accident events. The distribution of injuries shows that the launch phase and the general flying phase contained seminal events that led to the most severe accidents. From finer-grained analysis using the sub-phases described in figure 2, it was found that for the launch phase, the rotation into the climb was associated with most fatalities (3) and the recovery to speed after release was associated with most serious injuries (4). For the general flying phase most injuries occurred during ridge soaring (1 fatality, 2 serious and 5 minor injuries). The *search/descent* and *final glide* stages also had high numbers of injuries (3 serious and 4 minor).

Table 2
Totals, frequencies, and injuries categorized by where the seminal event was deemed to have taken place, in terms of the six higher level flight phases. Launches per accident are calculated from exposure data. The exposure data for pilots with 10 or fewer hours PIC was derived from the data published in Jarvis & Harris (2007).

High level Flight Phase	Injuries totals			10 or fewer hours experience			More than 10 hours	
	Fatal	Serious	Minor	Total accidents	Launches per accident	95% Confidence boundary	Total accidents	Launches per accident
1. Pre-flight	1	1	2	3	9,975	± 1,792	12	134,151
2. Launch	4	2	9	2	14,962	± 2,688	39	41,277
3. General flying	2	9	12	1	29,924	± 5,377	93	17,310
4. Circuit	1	0	5	3	9,194	± 1,792	68	23,674
5. Approach	0	3	12	11	2,720	± 489	61	26,390
6. Landing	0	1	4	8	3,741	± 672	58	27,755

For initial comparison of the two pilot experience groups, odds ratios were calculated between the groups for accident involvement in all 6 flight phases. An example of the cross-tab data for one test would be the total number of accidents with the seminal event in the landing phase against the total number of accidents without, all split by the two pilot groups. This was done six times; once for each of the flight phases. Only two phases resulted in significantly different odds between the two groups (i.e. where the 95% confidence interval did not include 1). These were the approach phase (Odds ratio: 2.864, in favour of inexperienced pilots having such accidents, with a 95% confidence interval from 1.28 to 6.42) and the general flying phase (Odds ratio: 10.55, in favour of experienced pilots having such accidents, with a 95% confidence interval from 1.41 to 78.765).

However, these results only show the odds of one type of accident occurring in relation to other types. They do not show the actual likelihood of these accidents occurring on a given flight. For this analysis, accident rates were required for both pilot groups, and all flight phases.

Accident rates were calculated for comparison of accident occurrence by flight phase seminal event between the two pilot groups. In the phases of flight where high numbers of accidents were observed, a more detailed analysis was performed. The rates across the six high-level flight phases (broken down by pilot experience) are shown in Table 2.

Whereas approach and landing were associated with the highest accident rates within the 10 or fewer hours group, the general flying and circuit phases had the highest accident rates within the experienced pilots' group. The general flying phase was the only phase where the accident rate for the over 10 hours group was higher than the 10 or fewer hours group (estimated at 1 launch in 17,310 against 1 in 29,924 for the 10 hours and fewer experience group). Fine-grained analysis of

the over 10 hours group showed that element 3c (search/descent) accounted for 40 seminal events in the general flying phase (43% of that category) by far the largest element. No seminal events appeared in this phase for the 10 or fewer hours experience group, and further analyses were not undertaken owing to the limited number of accidents.

The highest accident rate overall was associated with the inexperienced pilots and having a seminal event during the approach phase (1 in 2,720 launches by the mean estimate and one accident in every 3,209 launches by the lowest estimate). Even the lowest estimate is over eight times that of the more experienced group. The next highest was the landing phase with the lowest estimate of the accident rate at 1 in 4,413 launches, over six times higher than that for the over 10 hours experience group.

Discussion

The finding in this study that 80% of accidents were pilot related is only slightly higher than previous findings across the whole aviation domain, and is almost identical to the figure from US gliding activity reported by van Doorn & de Voogt (2007). Analysis of the primary causal categories demonstrates that accidents involving pilots with 10 or fewer hours flight experience were no different in their distribution of injury and aircraft damage to those of other, more experienced pilots. Analysis did not suggest that technical or external factors affected low experience pilots to a greater or lesser degree than more experienced pilots.

Forty-four percent of accidents were designated as having more than one *pilot induced* causal event; however, the narrative descriptions of events from accidents involving inexperienced pilots were more likely to involve multiple contributory causes.

Comparisons of flight phase findings in this study with previous research are slightly difficult as the definitions of phases used in this research are much more detailed than in previous studies and this analysis also uses seminal events (the first contributory factor leading to an accident – see Ingstad et al, 1990; Wiegmann et al, 2005) rather than categorizing on the basis of the phase of flight that the crash itself occurred. The General Flying phase in this study (phase 3 – see figure 2) most closely corresponds to the cruise phase used in previous research, although it is unclear in previous research whether cruise includes the circuit and approach phases of flight. The finding that most fatal accidents occurred during the cruise phase (van Doorn & Zijlstra, 2006) is not reproduced in this study since UK data shows that most fatal accidents occurred on Launch (see Table 2) despite this being the least frequent seminal accident phase after Pre-Flight. The discrepancy in these figures could be a result of the predominant use of aerotow launching in the US and the popular use of winch launching methods in the UK. However, the General Flying phase in this study (phase 3) did contain more seminal events than other categories and led to the largest number of injuries. This phase was particularly associated with the more experienced group of pilots (see table 2). Previous findings that Landing was the most frequent accident phase (e.g. van Doorn & de Voogt, 2007; O'Hare, 1994; AOPA, 2006) were supported by the current research only if seminal events taking place in the proceeding approach phase are included with those in the landing phase.

The significant difference in numbers of *out-landing* accidents between the two experience groups was to be expected considering that pilots with 10 or fewer hours solo experience would rarely fly out of range of the base airfield in the UK and therefore hardly ever need to make an out-landing. This explains the finding that the rate of accidents occurring to pilots with 10 or fewer hours experience was higher than for more experienced pilots, except when the seminal event occurred during the General Flying phase (table 2). In any cross-country, competition or soaring flight this phase makes up the majority of the duration of the flight and therefore experienced pilots carrying out such flights are exposed to this phase for much longer. Seminal events in flight phase 3c (Search/Descent) leading to accidents were exclusive to the more experienced group, accounting for 12% of all seminal events. This was partly because the seminal events resulting in these accidents involving a poor choice of landing area could usually be traced back to the initial field choice made during the descent and searching phase. Landings in unplanned locations such as farm pastures, crop fields, or scrubland bring dangers not associated with airfields, such as an uneven surface, obstacles and slopes, and these are usually smaller areas in which to land. It has previously been found that collisions with objects occurred mostly in terrain unsuitable for landing (van Doorn & de Voogt, 2007). These dangers are faced almost exclusively by pilots with over 10 hours experience who are more likely to undertake longer, cross-country flights than are novice pilots.

All the major flight phases except General Flying were found to have a much higher rate of seminal events leading to accidents for the 10 or fewer hours group compared with the group of pilots with more than 10 hours as PIC. The highest accident rate overall was associated with the pilots with 10 or fewer hours experience during the Approach phase (1 in every 2,507 launches using the mid-point estimate – see table 2). This was followed by the Landing phase (1 in 3,448 launches). Together, the Approach and Landing' phases made up 68% of the seminal events leading to accidents in the less experienced group. This supports previous findings for both gliders and airplanes regarding the frequency of landing accidents (O'Hare, 1994; van Doorn & de Voogt, 2007). The approach and landing are discrete phases of relatively short duration, but they both occur late in the flight and therefore have less opportunity for recovery. Glider pilots rarely have the option to go around and so must correct errors occurring in these stages with little time and little height remaining. This may challenge the ability of inexperienced pilots and therefore instructors may need to spend more time on these areas, particularly looking at recovery from events occurring during these phases. Of these two phases, the landing was associated with fewer injuries, which supports the finding that only 10% of landing accidents are fatal (van Doorn & de Voogt, 2007).

The Pre-Flight phase had a lower rate than both the Approach and Landing phases (one accident per 9,194 launches for the less experienced group – see table 2). This phase, however, showed the biggest discrepancy between 10 or fewer hours group of pilots and the more experienced group. Using the mid-point exposure estimates, the less experienced pilots were over 13.5 times more likely to have an accident when the seminal event occurred during the Pre-Flight phase. Pre-flight involves critical actions prior to take-off (particularly shutting the canopy and locking the airbrakes). It is possible that inexperienced pilots are making more

of these errors, but equally possible that the difference is caused by inexperienced pilots failing to recover from errors made at this stage, since a recovery from such an error would not feature in the accident statistics.

This research also demonstrates the limitations of applying odds ratios to accident data, as opposed to calculating accident rates based on exposure measures. For example, the odds ratio for the circuit phase was 2.16 (95% confidence interval: 0.63 – 7.35) suggesting that a more experienced pilot has over twice the odds of having an accident caused in that phase. However, the accident rates show that in fact inexperienced pilots have over twice the likelihood of having this kind of accident. This is because odds ratios of accident totals are based on the population of accident flights, rather than the population of all flights and therefore cannot be used to assess the likelihood of an accident occurring.

Conclusions

The distribution of seminal accident events among the various elements of a glider flight is quite different for inexperienced pilots than for pilots with more experience. Accident events within the 10 or fewer hours experience group were more likely to occur in the two major flight phases; Approach and Landing. These two phases made up 68% of seminal events leading to accidents in this group of pilots. The group of pilots with more than 10 hours of experience also had accidents originating in these phases but at a much lower rate. In general, the approach phase accounted for far more injuries than the landing phase. All flight phases other than General Flying showed higher accident rates for the pilots with 10 or fewer hours experience as PIC compared to more experienced pilots. Experienced pilots had a high rate of accidents originating during the General Flying phase, mostly in the Descent/Search sub-phase.

This research demonstrates the benefit of examining the flight phase in which the seminal event of an accident occurs, rather than classifying it by the flight phase in which the accident occurs. This allows better targeting of appropriate remedial actions. It also shows the benefits of analyzing accidents with respect to rates rather than frequency counts, as the type of flying undertaken by novice pilots differs considerably from that undertaken by more experienced pilots, especially when flying gliders. Simple frequency counts of accidents occurring in each flight phase can be misleading.

References

Aircraft Owners and Pilots Association (2006). *The Nall report 2006: Accident trends and factors for 2005*. Frederick, MD; The AOPA Air Safety Foundation.

Booze CF (1977). Epidemiological investigation of occupation, age, and exposure in general aviation accidents. *Aviation, Space & Environmental Medicine, 48,* 1081-91

British Gliding Association (2007). *UK glider accident database 1997 – 2006* (Unpublished internal document). Leicester, UK; British Gliding Association. 2007

British Gliding Association (2003). *BGA instructors' manual*, (2nd ed.). Leicester, UK: British Gliding Association.

Bureau of Air Safety Investigation (1996). *Human factors in fatal aircraft accidents.* Canberra, ACT; Department of Transport and Regional Development.

Feggetter, A.J. (1982). A method for investigating human factor aspects of aircraft accidents and incidents. *Ergonomics, 25,* 1065-1075

Gaur D. (2005). Human factors analysis and classification system applied to civil aircraft accidents in India. *Aviation, Space and Environmental Medicine, 76,* 501–5.

Hollnagel, E. (1998) *Cognitive reliability and error analysis method.* Oxford: Elsevier Science Ltd.

Ingstad, O., Rosness, R., Sten, T., Ulleberg, T., Rausand, M. & Lydersen, S. (1990). *Helicopter safety study. Main Report* (STF75 A90008). Trondheim, Norway: SINTEF.

Jarvis, S. & Harris D, (2007). Accident rates for novice glider pilots vs. pilots with experience. *Aviation, Space & Environmental Medicine, 78,* 1155-1158

Jenson RS. (1995). *Pilot judgement and crew resource management.* Ashgate, Aldershot, UK : Avebury Aviation

Kirwan, B & Ainsworth, L. (1992). *A guide to task analysis.* London: Taylor and Francis.

Lenne, M. & Ashby, K (2006). Characteristics of non-fatal general aviation crashes in Australia. *Human Factors & Aerospace Safety, 6,* 395-407

Li, G. (1994). Pilot-related factors in aircraft crashes: A review of epidemiological studies. *Aviation, Space & Environmental Medicine, 65,* 944-952.

Piggott, D. (1997). *Gliding. A handbook of soaring flight.* London, UK; A & C Black (Publishers) Ltd.

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

O'Hare, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37,* 1855 - 1869

Olsen S.O., & Rasmussen J. (1989). The reflective expert and the prenovice: notes on skill-, rule-, and knowledge-based performance in the setting of instruction and training. In: L. Bainbridge & S.A.R. Quintanilla (Eds.), *Developing skills with information technology* (pp. 9 – 33). London: Wiley.

Robson, C. (2002). *Real world research* (2nd ed.). Malden, UK: Blackwell Publishing.

Stewart, K (1994). *The glider pilot's manual.* Marlborough, UK; Airlife Publishing Ltd.

van Doorn, R.R.A. & Zijlstra, F. (2006). Epidemiological study of fatal and non-fatal glider accidents in the US, 2001-2005. *Human Factors & Aerospace Safety, 6,* 409-418.

van Doorn RRA & de Voogt AJ (2007). Glider accidents: An analysis of 143 cases, 2001-2005. *Aviation, Space & Environmental Medicine, 78,* 26-28.

Wiegmann, D., Faaborg, T., Boquet, A., Detwiler, C., Holcomb, K. & Shappell, S. (2005). *Human error and general aviation accidents: A comprehensive, fine-grained analysis using HFACS.* Final Report No: DOT/FAA/AM-05/24. Washington, DC; Office of Aerospace Medicine, Federal Aviation Administration.

Runway Incursions: An Industry Examination of FAA Initiatives and Objectives

William B. Rankin II
University of Central Missouri
TR Gaines 210
Warrensburg, MO 64093
wrankin@ucmo.edu
660-543-4455

Abstract

Previous research by Rankin in 1994 addressed the problem of runway incursions at the largest US towered airports and examined the perceptions of industry officials as to the effectiveness of the FAA initiatives or objectives implemented by the FAA Runway Incursion Plan of 1991. A similar study was completed in 2007 and investigates perception of industry officials as to the effectiveness of the FAA initiatives contained in the FAA Runway Safety Blueprint 2002-2004. For purposes of this paper, the studies are compared to see if there is a continued similarity of the perceived effectiveness by industry officials of the FAA initiatives or objectives. Since airport driver training was ranked as the number one initiate in the 1994 study and is not included in the FAA Runway Safety Blueprint 2002-2004, the 2007 study asked industry officials if airport driver training should, or should not be included in the FAA Runway Safety Blueprint.

Runway Incursions: An Industry Examination of FAA Initiatives and Objectives

The FAA (2008) just recently adopted the International Civil Aviation Organizations (ICAO) definition which states that a runway incursion is "Any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designated for the landing and takeoff of aircraft" (¶ 1). The National Airspace System (NAS) continues to experience approximately one runway incursion per week, which is classified as significant or a barely avoided collision (FAA 2004).

Runway incursions are divided into three classification types. These types include pilot deviations, operational deviations, and vehicle deviations. In the United States, pilot deviations account for approximately 57% of the total runway incursions, operational deviations account for 23%, and vehicle deviations account for 20% (FAA, 2004). After type, runway incursions are further stratified into four distinct categories by increasing severity, ranging from category D, the least severe, to category A, the most severe. Figure 1 illustrates the runway incursion categories by severity.

```
                        Increasing Severity
Low_____High
```

Category D	Category C	Category B	Category A
Little or no chance of collision but meets the definition of a runway incursion	Separation decreases but there is ample time and distance to avoid a potential collision	Separation decreases and there is a significant potential for collision	Separation decreases and participants take extreme action to narrowly avoid a collision, or there is a collision
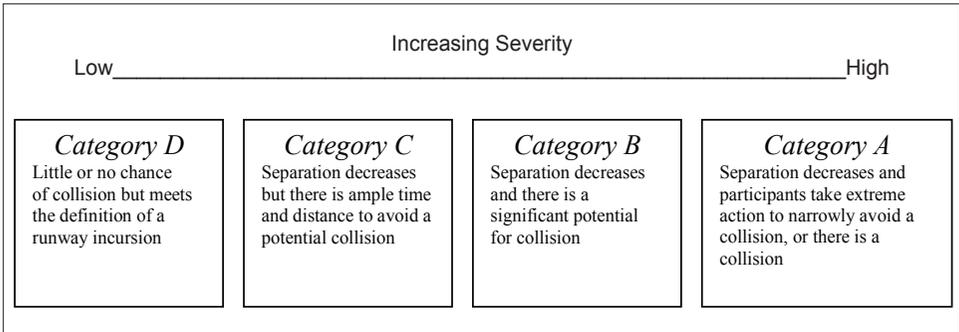
Figure1. Runway incursion categories by increasing severity. (FAA, 2004)

Although there has been a slight decrease in the overall number of deviations nationally, the number of category A and B vehicle deviations increased slightly from 2001. In 2001, there were 83 vehicle deviations, with five of the events being classified as category A and B incursions. In 2003, the number of vehicle deviations decreased to 60 events; however, nine of the events were classified as category A and B incursions (FAA, 2004). For the years 2000 through 2003, one of the most common errors that led to vehicle deviations was "Personnel or airport vehicles authorized on the movement area or airfield and instructed to hold short on the runway -- and whose operators verbally acknowledged the instructions -- entered the runway" (FAA, 2004, p. 26). The most likely causes for these types of occurrences included disorientation, loss of surface awareness, radio frequency congestion, and unfamiliar ATC language or procedures. Other causes may have included a lack of training on airport familiarity, airport layout, signs, and markings (FAA, 2004).

Aircraft repositioning is another common cause of vehicle deviations. Aircraft are often moved by *maintenance* taxi or *tug operations*. These terms refer to a mechanic (who is not a licensed pilot) taxiing or driving a service vehicle that is towing an aircraft on the airport surface. Airlines, charter operators, and air cargo operators often use these methods to reposition aircraft. If an incursion occurs during these types of operations on the airport surface, the FAA classifies the event as a vehicle deviation (FAA, 2004). From 2000 through 2003, there were 35 vehicle deviations involving airline maintenance taxi and tug violations (FAA, 2004). Of the 25 airports reporting the violations, the majority of these airports managed facilities with large-scale maintenance and cargo operations. Three of these vehicle deviations were classified as category B events (FAA, 2004).

The most serious runway incursion to date, a pilot deviation, occurred in Tenerife, Canary Island, on March 27, 1977, killing 583 people, and ranks as the worst disaster in aviation history (Clarke, 2002). According to Ragan (1997), the risk of misunderstanding ATC instructions communicated, particularly those between individuals from different cultures, via the radio is high and can have deadly consequences. Correctly understanding ATC information provided by the controller is essential for safe airport surface operations and can only be learned through a comprehensive training program. For example, a paper by Ragan (1997) concluded with an anecdote about a student pilot with limited English proficiency who

was asking the tower for permission to enter the traffic pattern to make a landing. The tower could not fully understand what he wanted, so the air traffic controller asked the student to state his intentions. The student responded by saying, "I intend to become a private pilot" (p. 34).

Goals, Objectives, Significance, and Limitations

The goals of this study were to evaluate the current *FAA Runway Safety Blueprint 2002-2004* objectives and to rank them in the order of their perceived effectiveness by industry officials. The study then compared the five most effective and the five least effective objectives obtained from current rankings with the results of similar research from the 1994 study.

The participants for this study were some of the largest US airports, major US domestic airlines, and the various aviation trade associations within the US aviation industry. A representative from each organization was chosen at random and designated to respond to a questionnaire. The participant list is attached as Appendix A.

Results from the 1994 study identified training of ground vehicle operators as the most effective FAA initiative to reduce runway incursions. Even though vehicle operators have traversed airport movement areas on a daily basis for many years, training of ground vehicle operators is conspicuously absent from mention in most literature. Also, the current runway safety objectives contained in the *FAA Runway Safety Blueprint 2002-2004* exclude any mention of airport driver training. As a result, participants were asked their opinion of the effectiveness of added ground vehicle operator training to the current *FAA Safety Blueprint* objectives.

This study is significant in that no previous study has been conducted to rank the *FAA Safety Blueprint 2002-2004* objectives by perceived effectiveness of US aviation industry officials or to see how the current objectives compared to a similar rankings of the FAA's *Runway Incursion Plan* initiatives in effect in the 1990s.

There were several potential limitations with respect to the survey instrument. These included (a) the effective sample size of participants, (b) the accuracy of the data provided by the participants, and (c) the statistical difficulty of using correlation to form conclusions. The 2007 survey instrument is attached as Appendix B.

Review of Literature

A review of literature examined training effectiveness from several aspects. A review of the history of FAA actions since the early 1990s lays the foundation for FAA initiatives and the development of the first action plan titled the FAA *Runway Incursion Plan.* A review of literature on the *Runway Safety Blueprint 2002-2004* addressed the primary causes for runway incursions and the complexities involved in solving runway incursions, as well as identifying the current FAA objectives to reduce the number of runway incursions.

According to the NTSB (1991) the foundation of the FAA's *Runway Incursion Plan* can be traced to the FAA's Assistant Administrator for Aviation Safety who in 1987 was directed by the Administrator to identify the causes of runway incursions and to formulate measures for alleviating this problem. The first phase of this effort resulted in the publishing of an Aviation Safety Bulletin and the creation of an informational video tape on runway incursions during March and June of 1988.

As a part of the second phase, a multidisciplinary team was formed under the overall direction of the Assistant Administrator for Aviation Safety. The team studied the problem and produced a report titled *Reducing Runway Incursions: An FAA Report.* The purpose of this report was to combine various perspectives on the runway incursion problem and to provide a basis for coordinating the efforts of the various FAA organizations into an integrated program for reducing runway incursions. The report stated:

> The team reviewed the various source materials related to runway incursions and talked to representatives of the user community (general aviation and commercial pilots, airport operators, and airport personnel), air traffic control personnel, and field personnel. Additionally, the team reviewed the ongoing problems and surveyed the activities of the agency – tasked with addressing these problems. Three recommendations resulted from the above efforts: (1) establish a steering committee on runway incursion reduction; (2) accelerate development and field deployment of the Airport Movement Area Safety System; and (3) emphasize the analysis of pilot-related causal factors in runway incursions. Specific recommendations were in five main areas as follows: (1) procedures in the cockpit and the control tower; (2) training of ground vehicle drivers; (3) awareness of the runway incursion problem; (4) signs, markings and lighting on airports; and (5) simplification of surface traffic movement. (cited in Rankin, 1994)

As a result of these recommendations, the FAA published a report titled *Runway Incursion Plan* in 1991. This report summarized ongoing actions and provided new initiatives intended to reduce runway incursions. The initiatives were summarized in a report by Harrison (1993), titled *Project Status – 1991 Runway Incursion Plan.* This report identified 45 initiatives that were eventually approved by the FAA for funding.

In a briefing paper titled *Runway Incursions and Surface Operations,* Harrison (1993) stated that these initiatives can be categorized into low and high technology. Some of the low technology initiatives identified were: (a) land and hold short warning lighting; (b) stop bar lighting, advisory circular on surface movement guidance and control; (c) airport diagrams (standard taxi charts); and (d) acknowledgement of hold clearances, and methods to ensure compliance. Examples of high technology initiatives include: (a) airport surface detection equipment; (b) runway status lights; (c) Airport Movement Area Safety System; (d) and an Airport Surface Traffic Automation System.

Surface movement safety, guidance, and control at major U.S. airports have continued as significant safety issues for almost two decades. Runway incursions occur when aircraft or vehicles travel onto a runway and conflict with aircraft cleared to take off or land on that same runway (FAA, 2004). In 1990, FAA Administrator Bussey wrote a letter addressed to the aviation industry on runway incursions stating:

One of the greatest areas of concern to not only the Federal Aviation Administration but the aviation community in general is the complexity of operations on or near active runways at controlled airports, and the hazards thus presented to passengers, crews, and aircraft. Infrequently, but more often than we would like to see, aircraft are involved in potentially catastrophic events during these operations that are referred to as runway incursions. (Rankin, 1994)

This same concern was voiced again by the FAA Administrator in the *Runway Safety Blueprint 2002-2004* transmittal letter, where Garvey stated:

One of my top priorities as Administrator has been the reduction of accidents and incidents caused by runway incursions. The *Runway Safety Blueprint 2002-2004* defines our strategy and prioritizes our efforts to reduce runway incursions. It presents the current state of runway safety at towered airports and identifies those areas where improvement is needed. (FAA, 2002a, ¶ 3-4)

The National Transportation Safety Board (NTSB) views the reduction of runway incursions as one of its most important transportation safety issues (FAA, 2004). The Department of Transportation (DOT) Office of Inspector General has identified runway incursions as one of the most difficult management challenges at the DOT. According to the FAA (2002b) recognition of the following key points is the basis for formulating and implementing solutions to improve runway safety for the nation:

1. Operational performance in the airport movement area must be further improved to reduce runway incursions.
2. Runway incursions are systemic, recurring events that are unintentional by-products of National Airspace System (NAS) operations.
3. Operations must be standardized to reduce risk at a time when growth is challenging runway infrastructure expansion.
4. Collision-avoidance safeguards need to be developed for the high-energy segment of runways, where aircraft accelerate for takeoff or decelerate after landing.
5. Human factors are the common denominator in every runway incursion. (FAA, 2002b, pp. 1-2)

The goals of the FAA *Runway Safety Blueprint* (2002a) are consistent with those identified by Rankin (1994) for all airports and include the following:

1. Develop and distribute runway safety education and training materials to controllers, pilots and all other airport users.
2. Increase surface safety awareness throughout the aviation community.
3. Assess and modify procedures to enhance runway safety.
4. Improve runway safety data collection, analysis, and dissemination.
5. Identify and implement enhancements to improve surface communications.
6. Increase situational awareness on the airport surface.

7. Support and deploy new technologies that reduce the potential for collision.
8. Implement site-specific runway safety solutions in coordination with local aviation communities. (p. 4)

According to Broderick (2008) the FAA is passing up a chance to implement a more comprehensive solution to the runway incursion problem as part of the FAA's National Airspace System modernization. In November 2007 the FAA issued a notice of proposed rulemaking (NPRM) which requires only Automatic Dependent Surveillance Broadcast (ADS-B) *out,* which provides basic aircraft information such as location and altitude. Adding ADS-B *in* to the FAA mandate would make other services possible, including a built-in way to transmit surface conflict warnings directly to pilots. Creating such a system, according to the NTSB, would be the most significant thing FAA could do to improve runway safety. Broderick (2008) further states that Vice Chairman Robert Sumwalt noted this in testimony submitted before the house runway safety hearing last February. To date, the FAA has not incorporated ADS-B *in* into the regulatory process.

Comments from the Airline Pilots Association (ALPA) – 1994 and 2007

Of particular interest in the 1994 study was a voluntary comment offered by Captain Mack Moore, Chairman of the Airports Standards Committee for the Air Line Pilots Association (ALPA), in response to the survey instrument. His comment on the effectiveness of the Airfield Smart Power initiative, under the broad area of Visual Aids, stated the following:

> Sequenced ground guidance light systems are the most promising of future systems for a number of reasons. First, they serve all aircraft on the airport and can do so in all weather conditions (as long as frozen contaminants do not cover the lights). Automated systems will eliminate many, but not all, human error opportunities. Automated ground guidance systems are viewed by the International Civil Aviation Organization (ICAO) as a capacity enhancing system for operations in all weather including excellent Visual Flight Rule (VFR) as well as a system that will prevent conflicts and incursions. (Captain Mack Moore, personal communication, February 1994)

Overall, in 1994 ALPA ranked the broad area of Visual Aids as the most important area in reducing the number of runway incursions-—closely followed by the broad areas of Technology second and Education third.

In a similar response to the 2007 survey, voluntary comments by Caption Robert Perkins, Chairman, ALPA Airport Ground Environment (AGE) Group, states:

> Objective 3 - We agree that the FAA has conducted research on surface operations memory aids, techniques, tools and training, regarding memory limitations, but there has also been no metric applied to determine its effectiveness. The results appear insignificant.
>
> Objective 7 – It is very important to implement a program for foreign air carrier pilot training, but we are not aware this has been implemented.
>
> Objective 9 – It is imperative to communicate runway safety concerns to all pilots, including international pilots as well.

Objective 21 – There has been very little produced and implemented as a result of the Phraseology Workgroup.

Objective 27 – The airport diagram issue is significant, as is the need to keep them current as construction occurs. This factor played a significant role in the LEX accident in 2006, where the NACO chart did not match the current layout. Airports must update their data in a timely manner.

Objective 33 – Moving map technologies are being allowed by the FAA, however, the standard to which they are held is detrimental to their deployment, although we concur that their accuracy is important. (Captin Robert Perkins, personal communications, March 25, 2008)

Methodology: Data Processing

This research was intended to be primarily a descriptive and non- experimental analysis that measured the perceived effectiveness of the FAA runway incursions initiatives and objectives by a survey of US aviation industry officials over a span of more than ten years.

For the purposes of this study, a combination of quantitative and limited qualitative methodologies using a five point Likert-type survey instrument was used (see Appendix B). This combination is called "Triangulation – A compatibility procedure designed to reconcile the two major methodologies by eclectically using elements from each of the major methodologies as these contribute to the solution of the major problem" (Leedy, 1993, p. 145).

Descriptive statistics was used to rank the five most effective and five least effective objectives outlined in the *FAA Runway Safety Blueprint 2002-2004* and compare them to the five most and least effective initiatives in the FAA's *Runway Incursion Plan* from 1991 to determine what disparities, if any, were apparent. FAA efforts to reduce runway incursions by the FAA to date have focused primarily on air traffic controllers and airline pilots, even though annual accidents at the 35 largest U.S. airports also involved airport vehicle deviations. Both survey instruments were prepared and distributed to all participants with a prepaid return envelope. Data collection from the latest survey was completed in the Spring of 2008.

Findings

Reliability of Survey Instrument

The fist step was to determine the internal consistency of the survey instrument. According to Norusis (2003):

In classical theory, a subjects' response to a particular item is the sum of two components: the true score and the error. The true score is the value of the underlying construct that is being measured; the error is the part of the response that is due to question-specific factors. The index most often used to quantify reliability is Cronbach's Alpha. Good scales have values larger than 0.8. (pp. 437-438)

In the case of this study, SPSS© software was used to calculate the Cronbach's Alpha value of 0.959 shown in Table 1 for the 36 survey questions used to study the *FAA Runway Safety Blueprint* in 2007.

Table 1
Cronbach's Alpha Reliability Statistics

Cronbach's Alpha	N of Items
.959	36

Descriptive Statistics Results

In the 1994 each participant was asked to rate the degree of effectiveness that each initiative in the *FAA's Runway Incursion Plan* had or will have on reducing the number of runway incursions using a five point Likert-type survey instrument where 0 represented the least effective and 5 the most effective. A 96% response rate was achieved in the 1994 study. The same type of survey instrument was used to collect data for the 2007 study. By contrast, however, nineteen of the 54 participants surveyed in the 2007 study responded achieving a 35% response rate. The mean for each initiative or objective was then determined using SPSS © software, which is the quotient of the sum of the values for each initiative or objective divided by the number of responses received for each initiative or objective. A comparison of the effectiveness of each initiative or objective was then determined by ranking each initiative or objective by its mean to establish the five most effective and the five least effective initiatives or objectives for both the 1994 and 2007 surveys. The survey results for the 1994 survey are contained in Figure 2 below.
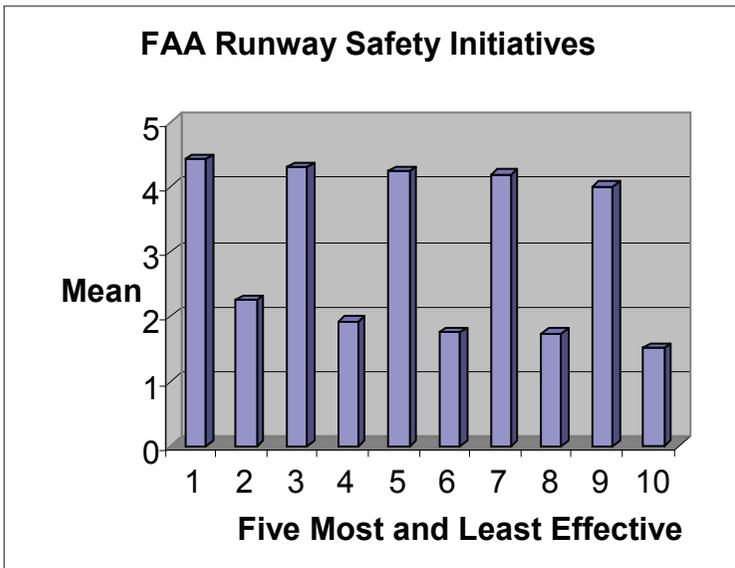

Figure 2. Rankings of Most and Least Effective Initiatives – 1994

In the 1994 survey the five most effective initiatives were identified by industry officials as: (a) Training of Ground Vehicle Operators with a mean value of 4.42; (b) Airport Surface Detection Equipment with a mean value of 4.30; (c) Stop Bar Lighting with a mean value of 4.23; (d) Airport Surface Traffic Automation with a mean value of 4.18; and (e) Airport Movement Area Safety System with a mean value of 4.00.

In the 1994 survey the five least effective initiatives were identified by industry officials as: (a) New Runway Safety Database with a mean value of 2.25; (b) Airport Technology Conference with a mean value of 1.92; (c) Audiotape on Runway Incursions with a mean value of 1.76; (d) Ground Movement Safety Awareness Products with a mean value of 1.75; and (e) New Computerized Database for Aircraft Performance with a mean value of 1.51.

The survey results for the participants responding to the 2007 survey are shown in Figure 3 below.
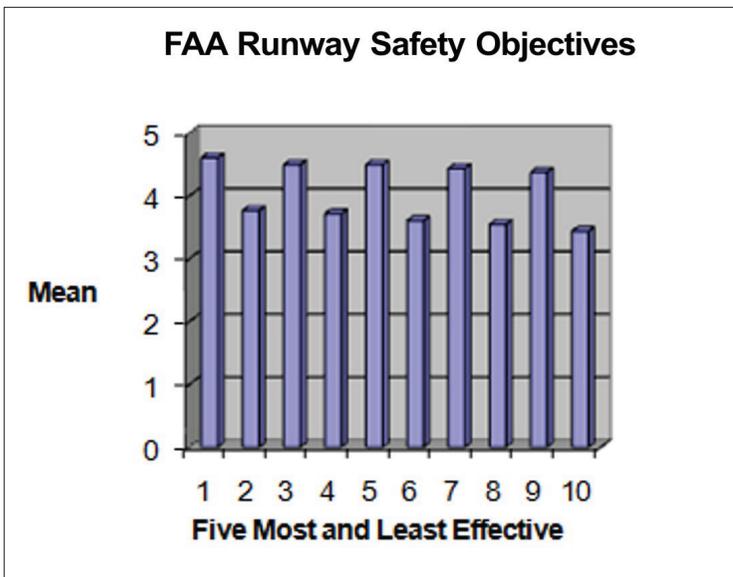

Figure 3. Rankings of Most and Least Effective Objectives – 1994

In the 2007 survey the five most effective objectives were identified by industry officials as: (a) Evaluate, and if appropriate, implement national procedures that require read backs of any clearance to enter a specific runway, hold short of a specific runway, or taxi into position and hold instructions with a mean value of 4.61; (b) Develop and evaluate a visual signal that provides direct warning to flight crews on final approach when the runway is occupied with a mean vale of 4.50; (c) Publish guidance on standard surface operations phraseology guidance for pilots and mechanics moving aircraft with a mean value of 4.44; (d) Assess selected Air Traffic procedures in terms of enhanced runway safety and recommend actions to retain, modify, or eliminate as appropriate with a mean value of

4.39; and (e) Improve runway safety data collection, storage, retrieval and distribution. Data and information useful for improving runway safety is contained in multiple databases operated by different organizations with a mean value of 4.33.

In the 2007 survey the five least effective objectives were identified by industry officials as:

(a) Create and accomplish a regional runway safety plan for each FAA region (every 18 to 36 months) tailored to specific operational and geographical needs with a mean value of 3.78.
(b) Improve the collection and analysis of operational error data by supporting the implementation and dissemination of the JANUS tool throughout the air traffic control environment with a mean value of 3.72.
(c) Maintain the published AMASS deployment waterfall schedule with a mean value of 3.61.
(d) Complete over 1,000 safety seminars per year incorporating runway safety, Runway Incursion Information Evaluation Program (RIIEP), surface movement Advisory Circulars and marking, signage and lighting as seminar themes with mean value of 3.56.
(e) Expand the role of Flight Service Station Specialists to provide runway safety information for towered and non-towered airports with mean value of 3.44.

Five of the participant's ranked objective 33 as the most effective; while two each stated that objectives 9, 17, and 22 were the most effective in reducing the number of runway incursions. As to the least effective objectives, two participants each stated that objectives 1, 25, 26, and 29 were the least effective in reducing the number of runway incursions. In response to the question - *In a 1994 survey on FAA objectives, airport movement area driver training ranked the most effective objective. Airport movement area driver training is no longer a specific objective. Should it be included as an FAA objective?* – Seventeen participants (89.5%) responded *yes,* while two (10.5%) responded *no.*

Discussion

Surface movement safety, guidance, and control at major US airports continues to be identified as significant safety issues for after two decades even with the best efforts put forth by the aviation industry and FAA. Aircraft or vehicles continue travel onto active runways and conflict with aircraft cleared to take off or land on that same runway on a daily basis. Moreover, the concerns expressed by FAA Administrator Bussey in the 1990s have once again been voiced by FAA Administrator Garvey in the Runway Safety Blueprint 2002-2004 transmittal letter.

FAA initiatives of the 1990s and the objectives of the Runway Safety Blueprint continue to focus in several main areas which include (a) procedures in the cockpit and the control tower; (b) awareness of the runway incursion problem; (c) signs, markings, and lighting on airports; as well as improvements in technology that assist in the simplification of surface traffic movement. However, several important areas have been overlooked by the FAA which need to be added in the Runway Safety Blueprint objectives' to adequately address the runway incursion problem. These include funding for ADS-B in to transmit surface conflict warning directly to

pilots and the training of ground vehicle operators, including training of aircraft tug and tow operators that re-position aircraft on a daily basis.

Finally, there is a significant difference in the range of effectiveness values noted in the 1994 survey versus the range noted in the 2007 survey. That difference may indicate that the FAA has been successful identifying objectives that are more effective in reducing runway incursions in the FAA Safety Blueprint of 2002-2004 as compared to the Runway Incursion Action Plan developed and implemented in the 1990s. The range of the 1994 study was from 4.42, the most effective initiative, to 1.51, the least effective initiative. The range of the 2007 study is from 4.61, the most effective initiative to 4.43, the least effective initiative. This may be an indication that the overall benefits of the objectives identified in the FAA Safety Blueprint of 2002-2004 are more effective overall in reducing the number of runway incursions than the initiatives adopted by FAA in the Runway Incursion Action Plan of the 1990s.

References

Broderick, S. (2008, April/May). Airports increase runway safety efforts. *Airport Magazine, 20*(2), pp. 47-49.

Clarke, B. (2002). *Runway incursions*. New York: McGraw-Hill.

Federal Aviation Administration (2008). *Runway safety.* Retrieved June 16, 2008, from http://www.faa.gov/runwaysafety/

Federal Aviation Administration (FAA). (2004, August). *FAA runway safety report FY 2000 - FY 2003*. Retrieved March 28, 2005, from www.faa.gov

Federal Aviation Administration (FAA). (2002a, July). *Runway safety blueprint 2002-2004*. Retrieved March 28, 2005, from www.faa.gov

Federal Aviation Administration (FAA). (2002b, June 21). *AC No: 150/5210-20 - Ground Vehicle Operations on Airports*. Retrieved April 12, 2005, from www.faa.gov

Harrison, M. J. (1993, January). Concepts for future airport surface movement guidance and control systems: Transitioning into the 21st century. *Research and Development Service Paper,* pp. 1-10.

Leedy, P. D. (1993). *Practical research: Planning and design* (5th ed.). New York: McMillan.

Norusis, M. J. (2003). *SPSS--Statistical Procedures Companion* (12.0 ed.). Upper Saddle River, NJ: Prentice Hall.

NTSB (1991, June). Aircraft Accident Report. *NTSB/AAR-91/05,* pp. 40-50.

Ragan, P. H. (1997, Winter). Aviation english: An introduction. *The Journal of Aviation/Aerospace Education and Research, 7*(2), 25-34.

Rankin, W. B. (1994). *Runway incursions: A censorious examination of runway incursions and the federal aviation administration's runway incursion program.* Unpublished master's thesis, Embry-Riddle Aeronautical University, Daytona Beach, Florida.

Appendix A

List of Participants
Airports:

Atlanta	Dallas Forth Worth
Los Angeles	Pheonix
Denver	Minneapolis
Las Vegas	Detroit
Houston – Intercontinental	Cleveland
Cincinnati	Philadelphia
Charlotte	Miami
Newark	Boston
St. Louis	Washington Dulles
Salt Lake City	Memphis
New York – LaGuardia, Newark, JFK	Seattle
San Francisco	Honolulu
Orlando	Baltimore
Chicago Midway	Fort Lauderdale
Portland	Cleveland
Washington Regan	Tampa
San Diego	

Airlines:

Air Alaska
Airtran
Aloha
American
America West
Continental
Delta
Frontier
Horizon
JetBlue
Midwest
Spirit
Southwest
United
US Airways

Aviation Trade Associations:

American Association of Airport Executives
Airports Council – North America
National Business Aircraft Association
Airline Pilots Association
Air Transport Association
Aircraft Owners and Pilots Association

Appendix B

FAA Runway Safety Blueprint Survey

Type of Organization Responding (circle one)   Airline   Airport   Industry Trade Association
Date _____

 Guidelines: Read each of the following objectives and circle the number that most appropriately represents your opinion. For the qualitative question on the final page, please clearly print your responses in the space provided.

The following objectives are effective:

Objective 1 – Complete over 1,000 safety seminars per year incorporating runway safety, RIIEP, surface movement Advisory Circulars and marking, signage and lighting as seminar themes.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 2 – Publish the airport vehicle surface operations Advisory Circular including best practices and SOP's. Coordinate with airport operators and associations to develop and ensure a successful implementation strategy and implementation plan.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 3 – Conduct research on surface operations memory aids, techniques, tools and training regarding memory limitations. Review existing course material to ensure that course curricula emphasize scanning techniques, anticipated separation and prioritization of control actions.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 4 – Require all tower controllers to complete approved training that emphasize steam effectiveness and situational awareness in an operational environment. In most cases, control tower functions are conducted by a team of controllers.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 5 –Develop course material and conduct training for initial and recurrent FAA Flight Standards Aviation Safety Inspectors (ASI) training. Enhance awareness of Certified Flight Instructors (CFI) and designated pilot examiners (DPE).
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 6 – Develop and implement enhanced tower controller training.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 7 – Implement a program for foreign air carrier pilot training.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 8 – Expand the role of Flight Service Station Specialists to provide runway safety information for towered and non-towered airports.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 9 - Publish a series of letters (two to four) to all pilots discussing runway safety.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 10 – Provide airport diagrams for towered airports via a link or other means as part of the standard Direct User Access Terminal Service (DUATS) to pilots.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 11 – Increase runway safety awareness within the aviation community by conducting at least one media emphasis project a year with trade and/or association periodical(s).

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 12 – Assess selected Air Traffic procedures in terms of enhanced runway safety and recommend actions to retain, modify, or eliminate as appropriate.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 13 – Implement national standardized requirements for tower positions to ensure uniform, effective, and sustained situational awareness practices relating to surface operations.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 14 – Implement standardization of national equipment and procedures for Runway Incursion Devices (RID).

   1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 15 – Publish the best practices/SOP appendix to each of the two pilot Surface Movement Advisory Circulars (AC 120-74 and 91-73) and widely disseminate to the General Aviation Air Carrier communities.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 16 – During enroute inspections, Aviation Safety Inspectors ensure that pilots have current surface movement charts available and that they are in use.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 17 – Develop advisory circulars that address procedures, best practices, and SOPsfor airline maintenance taxi operators and for tug and tow vehicles while operating on the airport surface. Include "best practices" and a checklist. Coordinate with industry (airport managers, airlines, and Fixed Base Operators (FBO)) to use this information in training courses.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 18 - Disseminate and provide training for the Runway Incursion Information Evaluation Program (RIIEP) to all FAA Safety Inspectors. Develop data collection and analysis system that provides report and trend information regarding runway incursions caused by pilot deviations that can be used by safety inspectors, flight instructors, examiners and others.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 19 - Improve runway safety data collection, storage, retrieval and distribution. Data and information useful for improving runway safety is contained in multiple data bases operated by different organizations.

   1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 20 – Improve the collection and analysis of operational error data by supporting the implementation and dissemination of the JANUS tool throughout the air traffic control environment.

  1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 21 – Complete and publish results from Phraseology Workgroup.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 22 – Evaluate, and if appropriate, implement national procedures that require read backs of any clearance to enter a specific runway, hold short of a specific runway, or taxi into position and hold instructions.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 23 – Publish guidance on standard surface operations phraseology guidance for pilots and mechanics moving aircraft.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 24 – Issue guidance on vehicle operations near active runways.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 25 - Complete the airport paint marking study and revise standards in the advisory circular, if appropriate based on a review of study results by the Office of Airport Safety and Standards.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 26 - Complete airport design and operations study and develop airport configuration and operational procedures enhancements in sufficient detail to evaluate at least one airport in an operational environment.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 27 – Ensure towered airports have current airport diagrams by the end of December 2003 and they are available in Government publications.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 28 – Maintain the published AMASS deployment waterfall schedule.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 29 – Develop high-level requirements for Runway Status Lights (RWSL) and validate alternative implementation methods through conduct of field demonstrations.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 30 – Conduct evaluations of existing low-cost technologies.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 31 – Meet published ASDE-X milestones.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 32 – Evaluate moving map technologies in an operational environment – using either aircraft or surface vehicles.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 33 – Develop and evaluate a visual signal that provides direct warning to flight crews on final approach when the runway is occupied.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 34 – Develop a surface "road map" for a low-cost technology architecture and periodically release Broad Agency Announcements (BAA) to solicit industry surface safety technology ideas and concepts.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 35 - Create and accomplish a regional runway safety plan for each FAA region (every 18 to 36 months) tailored to specific operational and geographical needs.

1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

Objective 36 – Implement an aggressive runway safety "special emphasis" at selected airports that results in reducing runway incursions.
1 – Strongly Disagree 2 – Disagree 3 – No Response 4 – Agree 5 – Strongly Agree

General Questions

In your opinion, which of the above objectives is the most effective _____

In your opinion, which of the above objectives is the least effective _____

In a 1994 survey on FAA objectives, airport movement area driver training ranked the most effective object. Airport movement area driver training is no longer a specific objective. Should it be included as an FAA objective? **Yes   No**

Other Comments: _____
_____
_____

Implementing Likelihood Alarm Technology in Integrated Aviation Displays for Enhancing Decision-Making: A Two-Stage Signal Detection Modeling Approach

Ernesto A. Bustamante
University of Idaho
Psychology & Communication Studies
Student Health Centre Room 006
Moscow, ID 83844-3043
Phone: (208) 885-0338
Fax: (208) 885-7710
E-mail: ernestob@uidaho.edu

Abstract

Decision-making plays a crucial role in efficient alarm response, and, consequently, in adequate system-monitoring performance. The decision-making process involved while interacting with integrated aviation displays can be analyzed using an information-processing model, involving different stages of processing. The following manuscript presents a two-stage signal detection modeling approach of decision making while interacting with integrated aviation displays that allows researchers to partition these separate processing stages. Two experiments were conducted, examining the effects of likelihood alarm technology (LAT), workload, and task-critical information (TCI), on decision-making accuracy and bias. It was hypothesized that these factors would have differential direct and indirect effects on decision-making accuracy and bias during different stages of processing. Results were consistent with the postulated hypotheses as well as with prior research. The findings from this research effort demonstrated the superior advantage of using LAT to increase decision-making accuracy, decrease decision-making bias, and ultimately enhance monitoring performance.

Implementing Likelihood Alarm Technology in Integrated Aviation Displays for Enhancing Decision-Making: A Two-Stage Signal Detection Modeling Approach

Technological advances have made the use of integrated displays a common practice in aviation (Bliss, 2003) and air traffic control (Masalonis & Parasuraman, 2003). The increased use of integrated displays has changed the role of pilots from aircraft operators to system monitors (Parasuraman & Riley, 1997). Human monitors are notoriously ineffective in complex situations characterized by high levels of workload (Woods, 1995). Consequently, engineers and designers have

developed alarm systems to assist human monitors (Papadopoulos & McDermid, 2001). Advanced sensor technologies and fault-diagnosis algorithms have allowed alarm systems to detect the presence of dangerous conditions effectively (Tumer & Bajwa, 1999).

Alarm Systems, the Cry-Wolf Effect, and the Attentional Capture Effect

The primary purposes of alarm systems are to detect dangerous conditions and attract pilots' attention so that they can either avoid or escape problems (Xiao & Seagull, 1999). Ideally, systems should issue alarms only when there is an actual underlying problem present. However, because of legal implications, system designers tend to follow the "engineering fail-safe approach," setting the threshold of alarm systems low enough to alert pilots of even the slightest possibility of a problem (Swets, 1992). Moreover, the rare occurrence of dangerous conditions makes it difficult for designers to develop alarm systems that emit a low number of false alarms (Parasuraman, Hancock, & Olofinboba, 1997). Consequently, most alarm systems generate many false alarms (Getty, Swets, Pickett, & Gonthier, 1995). Frequent false alarms may cause a loss of trust in the system and reduce pilots' compliance with alarm signals (Breznitz, 1983). Continuous exposure to frequent false alarms can also create a desensitization effect to the extent that pilots may become accustomed to alarm signals, and, as a result, such signals may fail to produce the necessary attentional capture effect. As a result of this cry-wolf effect and potential lack of attentional capture effect, pilots often ignore or cancel alarms without searching for additional information that could help them detect the presence of dangerous conditions (Sorkin, 1988).

One of the problems associated with the cry-wolf effect is that most alarm systems use binary alarm technology (BAT), issuing a single type of alarm once conditions exceed a predetermined threshold, regardless of the underlying nature of the problem (Woods, 1995). Therefore, it is often difficult for pilots to differentiate false alarms from true signals. Another problem that is related to the cry-wolf effect is that most complex alarm systems use an integrated display. According to the proximity compatibility principle (Wickens & Carswell, 1995), integrated information tends to be more effective than information presented through different displays when tasks require integration of different sources of information (O'Brien and Wickens, 1997). The proximity compatibility principle has advantages and disadvantages. Tasks that are closely related benefit from integrated displays. Display integration eliminates pilots' need to divide their attention among different sources of information (Wickens & Liu, 1988). However, this creates the extra burden on pilots of having to navigate through different layers within the same display.

For example, in the field of commercial aviation, due to the introduction of the "glass cockpit," critical information about system status has been highly integrated within a small number of primary display units, such as the Engine Indication and Crew Alert System (EICAS). As a result of this integration, not all critical information is readily available to pilots for them to determine whether there actually is a problem when they encounter an alarm signal. Instead, pilots frequently must "go head down" to look for this information by searching through different display pages.

These problems are accentuated as workload increases because pilots often need to allocate their time and effort among different tasks (Maltz & Meyer, 2001). Under low-workload conditions, attending to alarms may not degrade performance to a great extent. Therefore, pilots may tend to respond more often to alarm signals under such conditions, even if they are usually false (Bliss & Dunn, 2000). However, during high-workload conditions, attending to alarms impose a greater cost on performance. Consequently, pilots may over-comply with alarms if they are likely to be valid or ignore them completely if they are likely to be false (Bliss, 2003).

One possible solution to the cry-wolf effect is to provide pilots with task-critical information (TCI) to help them make better decisions when they encounter alarm signals. The problem with this approach is that under periods of high workload, pilots may not have enough attentional resources to process such information. A more effective way to mitigate the cry-wolf effect may be to use Likelihood Alarm Technology (LAT) (Bustamante, 2007). Likelihood alarm systems (LASs) issue different types of signals depending on the likelihood that an underlying problem actually exists (Sorkin, Kantowitz, & Kantowitz, 1988). The purpose of using LASs is to allow pilots to make better decisions regarding which signals they can more likely ignore and to which alarms they need to immediately respond.

Alarm Systems and Decision Making
Decision-making plays a crucial role in responding to alarms, particularly under varying levels of workload and in the presence of different sources of information. When pilots encounter alarm signals, they have to perform a series of alarm-initiated activities, which include, among others, analyzing and investigating the nature of the alarms and correct the underlying problem (Stanton & Baber, 1995). The decision-making process involved in alarm response can be analyzed using an information-processing model, involving different stages of processing. In its simplest form, information processing involves a series of stages, including perception, attention, decision making, and response execution (Wickens, 1987). Alarm signals are typically designed to be salient enough for pilots to perceive them (Edworthy & Stanton, 1995). Therefore, the perceptual stage does not play a very important role in alarm response. Attention and decision making, on the other hand, play a crucial role in fault diagnosis and the detection of system malfunctions (Moray, 1981). Last, the response execution process itself is mostly a result of the decision-making process (Wickens, 1987).

Sorkin and Woods (1985) characterized the response process using a two-detector signal detection model. The underlying idea behind this model is that tasks that require pilots to interact with alarm systems resemble a signal detection task in which pilots constantly monitor a variety of sources of information for specific problems and make decisions based on the information they obtain and the outputs generated by the alarm system. The main advantage of using this model is that it allows researchers to partition the human response process by examining variables that affect pilots' bias to respond to alarm signals separate from their ability to detect problems. However, most previous research has focused on the use of alarm systems to aid performance on a detection task

where critical information about the task has been readily available (Lehto & Papastavrou, 1998; Maltz & Meyer, 2001; Meyer, 2001; Robinson & Sorkin, 1985; Sorkin, Kantowitz, & Kantowitz, 1988).

A potential limitation of this two-detector approach is that it does not allow researchers to examine how pilots weigh outputs provided by the system apart from the information they obtain directly from other systems and the environment. Furthermore, in applied settings, due to display integration, pilots do not always have direct access to system-status information. Another problem with this approach is that it does not allow researchers to examine which level of information processing (i.e., attention or decision-making) is most affected by the system's parameters.

A Two-Stage Signal Detection Modeling Approach
The following manuscripts presents a two-stage signal detection modeling approach of decision making while interacting with integrated alarm systems that allows researchers to partition these two separate information processing stages (see Figure 1).
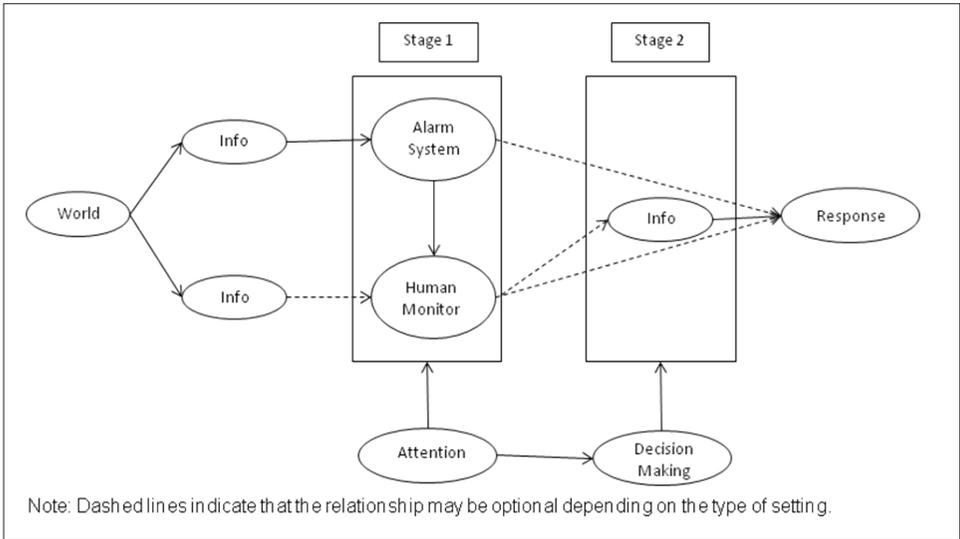

Note: Dashed lines indicate that the relationship may be optional depending on the type of setting.

Figure 1. Two-stage signal detection model of decision-making.

This model is based not only on the theoretical foundation of information processing stages (Wickens, 1987), but also on the practical implications that result from the high level of display integration used in many applied settings. Furthermore, this model is based on an adaptation of the *a-b* Signal Detection Theory Model (Bustamante, 2008) to Sorkin and Woods (1985)'s signal detection analysis of systems with human monitors.

When pilots encounter alarm signals, they typically analyze input data, diagnose the underlying nature of the problem, and make a corrective or evasive action. Due to display integration, when an alarm is perceived, pilots may decide to either ignore it or search for further information. This constitutes the first stage of the decision making process, which is influenced mostly by attentional capacity. If

pilots decide to search for further information, they have to diagnose the nature of the underlying problem and make a corrective or evasive decision. This constitutes the second stage of the decision-making process, which is influenced mostly by information processing and focuses mostly on making a decision and implementing it. Additionally, depending on the level of automation of the alarm system and the level of self-confidence of pilots, either component of the human-machine system may bypass one or more of the information processing stages and make responses automatically.

Goal of this Research

The purpose of this research was to examine the effects of implementing LAT in integrated aviation displays on the two stages of decision making involved in monitoring tasks during varying workload levels and in the absence and presence of TCI. Research suggests that the combined performance of the human-machine system in an alerted monitoring task is moderated by workload and the presence of TCI, and limited by the system's parameters and how operators adjust to such parameters (Pollack & Madans, 1964).

Likelihood Alarm Technology (LAT)

Research suggests that LAT affects pilots' compliance with warnings (Wogalter, Young, Brelsfor, & Barlow, 1999). Pilots adjust their decision-making bias according to the system's threshold (Maltz & Meyer, 2001). This idea is consistent with the phenomenon of probability matching, which suggests that pilots adjust their response rate according to the system's reliability (Bliss, Gilson, & Deaton, 1995). LASs issue different types of signals by having different predetermined thresholds. Once conditions exceed the low threshold, LASs issue low-likelihood alarms. Once conditions exceed a higher threshold, LASs issue high-likelihood alarms. The purpose of LAT is to decrease pilots' response bias towards low-likelihood alarms and increase their response bias towards high-likelihood alarms. As a result, pilots may respond more often to true signals and ignore false alarms. This may increase their decision-making accuracy during the first processing stage, and, as a result, increase their decision-making accuracy during the second processing stage.

Workload

Workload may have a direct effect on pilots' decision-making bias during the first processing stage (Maltz & Meyer, 2001), making them less likely to respond to alarm signals regardless of whether they are true signals or false alarms (Bliss & Dunn, 2000). As a result, workload may have an indirect effect on pilots' decision-making bias during the second processing stage because of its dependency on the first processing stage (Wickens, 1987). Therefore, if pilots are less likely to search for further information when they encounter alarm signals, they may, in turn, be less likely to make necessary corrective or evasive decisions. Workload may also have a direct impact on pilots' decision-making accuracy during the second processing stage because it may limit the amount of resources applied to processing and diagnosing system-status information (Sorkin & Woods,1985).

Task-Critical Information (TCI)

TCI may increase pilots' decision-making accuracy during the first processing stage by serving as a guide to differentiate true signals from false alarms (Bliss, Jeans, & Prioux, 1996). Consequently, TCI may enhance pilots' decision-making accuracy during the second processing stage. However, research suggests that TCI may also influence pilots' decision-making bias during the first stage, making them less likely to search for further information (Bliss, Jeans, & Prioux, 1996). As a result, TCI may have an indirect impact on pilots' decision-making bias during the second processing stage, making them less likely to take necessary corrective actions.

Another point to consider about TCI is the degree of redundancy with the alarm system. High levels of dependency between redundant sources of information are common in most applied settings. Research suggests that the advantage of TCI decreases with increased dependency of the information sampled by the alarm system (Elvers, 1997). The more the two sources of information are related, the fewer benefits TCI provides to pilots interacting with an alarm system.

Hypotheses

LAT. It was hypothesized that LAT would have a direct impact on decision-making accuracy during the first processing stage and, as a result, an indirect effect on decision-making accuracy during the second processing stage (Sorkin, et al. 1988). Figure 2 shows a schematic representation of these hypotheses.
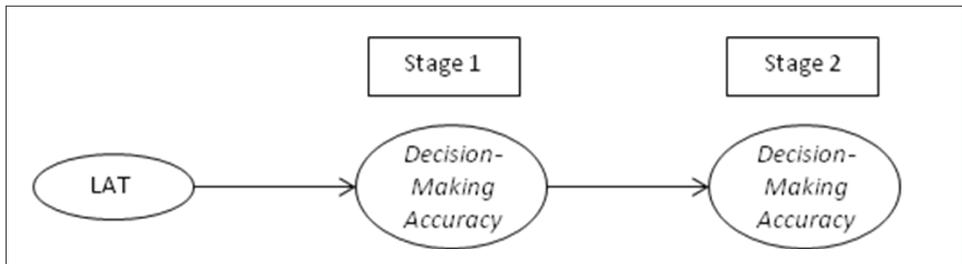

Figure 2. Schematic representation of likelihood alarm technology hypotheses.

Workload. Workload was expected to have a direct effect on decision-making bias during the first processing stage and an indirect effect on decision-making bias and accuracy during the second processing stage (Bliss & Dunn, 2001, Maltz & Meyer, 2001). Workload was also expected to have a direct effect on decision-making accuracy during the second processing stage (Sorkin & Woods, 1985). Figure 3 shows a schematic representation of these hypotheses.
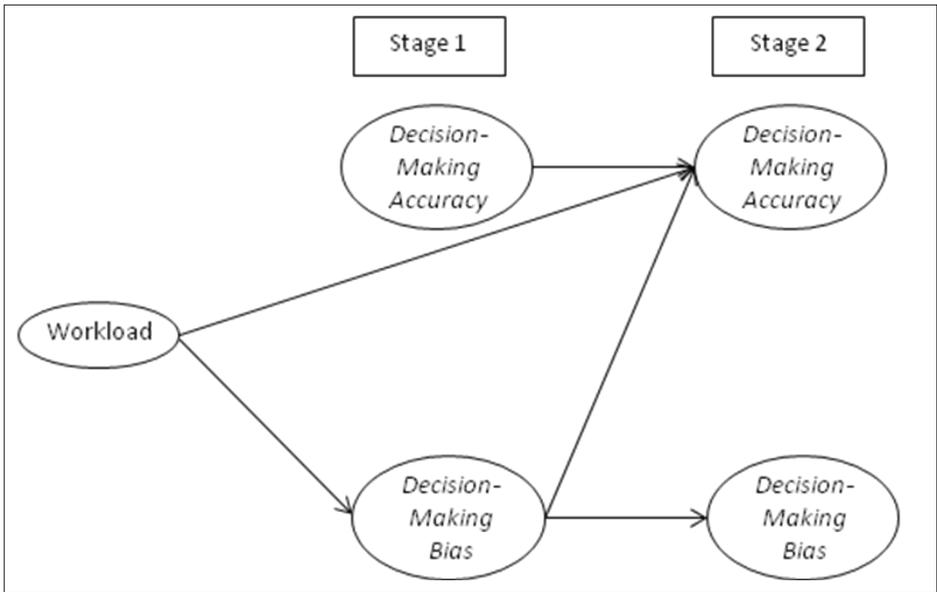
Figure 3. Schematic representation of workload hypotheses.

TCI. It was hypothesized that TCI would have a direct effect on decision-making accuracy during the first processing stage (Bliss et al., 1996), moderated by the type of system (Elvers, 1997). Because of the level of dependency between TCI and the LAT used in this particular research, the effect of TCI on decision-making accuracy was expected while operating with a traditional binary alarm system (BAS) only. It was also expected that TCI would have a direct effect on decision-making bias during the first processing stage, such that overall alarm responsiveness would be lower in the presence of TCI (Bliss et al., 1996). Figure 4 shows a schematic representation of these hypotheses.
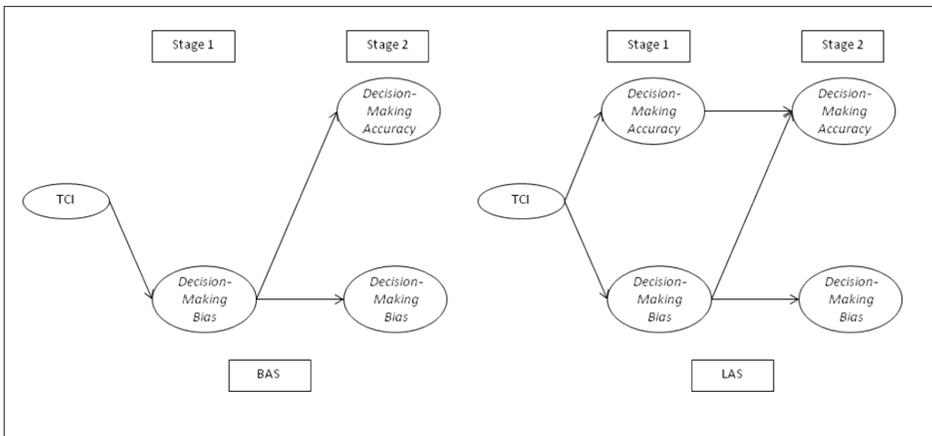

Figure 4. Schematic representation of task-critical information hypotheses.

Method

Experimental Design

A 2 x 2 repeated-measures design was used for Experiments 1 and 2. System (BAS, LAS) and workload (low, high) were manipulated within groups. The only difference between the two experiments was the presence of TCI in Experiment 2. A 2 x 2 x 2 mixed design was used to analyze the combined data from both experiments. Therefore, TCI (no, yes) was manipulated between groups. The data collection protocols for both studies were previously approved by the Institutional Review Board and were in accordance with the APA ethical standards.

Participants

Fifty-four (13 males, 41 females) university students participated in this study. Thirty students participated in Experiment 1, and 24 students participated in Experiment 2. Participants ranged from 18 to 42 years of age (M = 22.70, SD = 5.58), and they all had normal or corrected-to-normal vision and hearing.

It is important to address the disproportional gender distribution of this sample. This is a typical problem that researchers encounter when they use a participant pool of psychology students, who are predominantly females. This issue could be problematic to the extent that gender differences could systematically account for variations in the findings. However, to the author's knowledge, neither theory nor prior empirical findings suggest that gender is a critical factor that affects decision making while interacting with decision support tools.

Materials and Apparatus

Multi-Attribute Task Battery (MATB). The MATB (see Figure 5) is a computer program that was developed to assess human performance and workload under different conditions for several research purposes (Comstock & Arnegard, 1992). For this research effort, the dual-axis compensatory-tracking and resource-management sub-tasks were used to simulate the primary flight tasks. The MATB monitoring task was used to present TCI in Experiment 2.
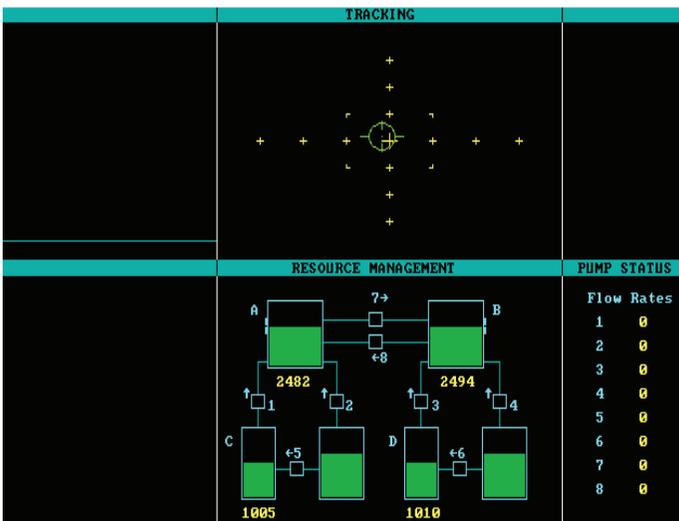

Figure 5. Multi-Attribute Task Battery.

Dual-Axis Compensatory-Tracking Task. The main purpose of this task was to simulate the key function that pilots need to perform to fly an airplane, which is to maintain level flight. Participants' job was to keep a circle as close to the center as possible using a dual-axis joystick.

Resource-Management Task. The main purpose of this task was to simulate another important function that pilots need to perform as they fly an airplane, which is to make sure that they have an optimal level of fuel in their tanks. Participants' job was to keep an optimal level of fuel on the two main tanks, while preventing any of the secondary tanks from being depleted.

Secondary Engine-Monitoring Task

The main purpose of this task was to simulate a crucial secondary function that pilots need to perform to maintain flight safety, which is to ensure that they have at least one fully functioning engine at all times. Participants performed this task with the aid of a simulated EICAS (see Figure 6).

Figure 6. Simulated engine indicating and crew alerting system

Participants were tasked with either acknowledging or ignoring alarms emitted by EICAS. This step constituted the first processing stage. Each alarm was composed of a visual stimulus as well as an auditory stimulus, which was presented

to participants through a set of sound-attenuated headphones. Given that the average ambient noise level was 55 dB(A), the auditory stimulus was presented at 65 dB(A) to ensure that the alarms were salient enough for participants to notice them. In case participants decided to acknowledge a particular alarm, they gained access to additional system-status information (see Figure 7) to help them make a corrective action when necessary. This step constituted the second processing stage.
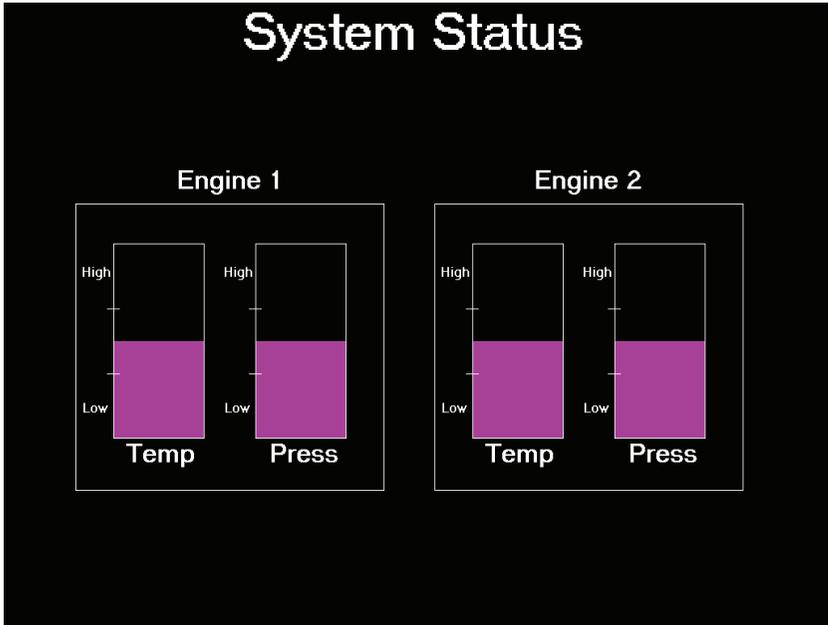

Figure 7. System-status information

Alarm System
   Binary Alarm System (BAS). The performance of the BAS was modeled based on prior research (Bustamante, Anderson, & Bliss, 2004). The overall reliability of this system was 18%, which meant that 18 out of every 100 alarms were true. This system always generated one type of alarm signal. Therefore, the likelihood of every signal was 18%. The visual component of the alarm signal consisted of a yellow square accompanied by the word "WARNING." The auditory component of the alarm signal consisted of a simple sine wave of 500 Hz, presented at 65 dB(A). The ambient sound pressure level was approximately 45dB(A).

   Likelihood Alarm System (LAS). The overall performance of the LAS was the same as the BAS. However, the LAS generated two types of alarm signals that differed in their likelihood to be true. The low-likelihood signals were 5% likely to be true, and they consisted of the same stimuli used for the binary system. The high-likelihood signals were 88% likely to be true. The visual component of the high-likelihood alarms consisted of a red circle accompanied by the word "DANGER." The auditory component of the high-likelihood signals consisted of a simple sine wave of 2500 Hz, also presented at 65dB(A).

Task-Critical Information

Depending on the experiment, participants received TCI using the system-monitoring display of the MATB. This display consists of four gauges that indicate temperature and pressure levels of two engines. Different fluctuation patterns were indicative of potential engine malfunctions with varying degree of likelihood. When only one gauge fluctuated outside of the normal range, there was a 5% likelihood of an engine malfunction. When two of the gauges fluctuated out of range consecutively, there was an 88% likelihood of an engine malfunction. Fluctuations were always accompanied by an alarm signal, and the likelihood of the signal was in perfect agreement with the type of fluctuation pattern.

Workload

Workload was manipulated by automating the dual-axis compensatory-tracking task and by introducing a series of random pump malfunctions in the resource-management task.

Dependent Measures

Decision-making accuracy and bias were derived using the *a-b* SDT Model (Bustamante, 2008). The *a-b* SDT model is based on the work of Bustamante (2008), who offered alternative measures of decision-making accuracy (*a*) and response bias (*b*) that do not rely on the underlying assumptions of the traditional SDT model. Instead, *a* and *b* are based simply on the outcome matrix, defined by the proportion of hits, false alarms, misses, and correct rejections. Within the *a-b* SDT model, decision-making accuracy is conceptually defined as the tendency to make correct responses (i.e., hits and correct rejections). Response bias, on the other hand, is conceptually defined as the tendency to make affirmative responses (i.e., hits and false alarms).

In most applied settings, researchers are concerned with the ability of humans and automated systems to make accurate decisions. Therefore, Bustamante (2008) first replaced the term "sensitivity" with "accuracy" (*a*) and defined it as the weighted sum of the proportion of correct affirmative and negative responses, or

$$\alpha = .5 * p(HI) + .5 * p(CR) \qquad (1)$$

where,

a = accuracy
p(HI) = proportion of hits
p(CR) = proportion of correct rejections

Bustamante (2008) defined response bias (*b*) as the weighted sum of the proportion of correct and incorrect affirmative responses, or

$$b = .5 * p(HI) + .5 * p(FA) \qquad (2)$$

where,

b = response bias
p(HI)= proportion of hits
p(FA) = proportion of false alarms

One of the many advantages of the *a-b* SDT Model is that the alternative *a* and *b* measures may be interpreted more intuitively. With regard to *a*, a score of 0 indicates the complete lack of ability to make accurate decisions. A score of .5 indicates performance at chance level, and a score of 1 indicates perfect decision-making accuracy. With regard to *b*, a score of 0 indicates a lack of affirmative responsiveness. A score of .5 indicates an unbiased level of responsiveness, and a score of 1 indicates a complete response bias toward affirmative responses.

Procedure
As part of this study, participants completed four experimental sessions, which varied according to the level of workload and the type of alarm system. The order in which participants completed these sessions was randomized to minimize carryover effects. Participants came to the lab one at a time and at different times. When they came into the lab, they first read and signed an informed consent form. Then, experimenters explained to them the nature of the study and provided them with the instructions on how to perform the tasks. Next, participants performed a series of practice sessions. Before each experimental session, experimenters informed participants of the overall reliability of the system and the likelihood of each type of alarm. Then, participants performed each experimental session, which lasted a total of 30 min. After they finished the second experimental session, participants took a 5-min break. After the last session, experimenters debriefed participants and thanked them for their participation.

Results

Experiment 1
Two 2 x 2 repeated-measures ANOVAs were used to analyze decision-making accuracy and bias during the first processing stage. System and workload were used as independent variables. Decision-making accuracy and bias were used as dependent measures. Results showed a significant main effect of system on decision-making accuracy, $F(1, 29) = 145.37$, $p < .001$, partial $\eta^2 = .83$. Participants' decision-making accuracy was significantly higher when they used the LAS ($M = .77$, $SE = .01$) rather than the BAS ($M = .51$, $SE = .01$). Results also showed a significant main effect of workload on decision-making bias, $F(1, 29) = 45.76$, $p < .001$, partial $\eta^2 = .61$. Participants' decision-making bias was significantly lower during high workload ($M = .49$, $SE = .01$) than during low workload ($M = .60$, $SE = .01$).

A 2 x 2 repeated-measures ANCOVA was used to analyze decision-making accuracy and bias during the second processing stage. System and workload were used as independent variables. Decision-making accuracy and bias during the first processing stage were used as mediating variables. Decision-making accuracy and bias during the second stage were used as the dependent measures. Results showed a significant mediating effect of decision-making accuracy during the first processing stage on decision-making accuracy, $F(1, 27) = 8.18$, $p < .01$, partial $\eta^2 = .23$, and bias, $F(1, 27) = 20.07$, $p < .001$, partial $\eta^2 = .43$ during the second processing stage. As participants' decision-making accuracy during the first processing stage increased, their decision-making accuracy during the second stage also increased ($r = .49$, $p < .001$), as well as their decision-making bias ($r = .49$, $p < .001$). Last, results showed a significant mediating effect of decision-

making bias during the first processing stage on decision-making accuracy, $F(1, 27) = 8.18$, $p < .001$, partial $\eta^2 = .23$, and bias, $F(1, 27) = 20.07$, $p < .001$, partial $\eta^2 = .43$ during the second processing stage. As participants' decision-making bias during the first processing stage increased, their decision-making accuracy during the second processing stage increased ($r = .75$, $p < .001$), as did their decision-making bias ($r = .75$, $p < .001$). Figure 8 shows a schematic representation of these results.

*Mean difference between LAT and BAT

**Mean difference between low and high workload
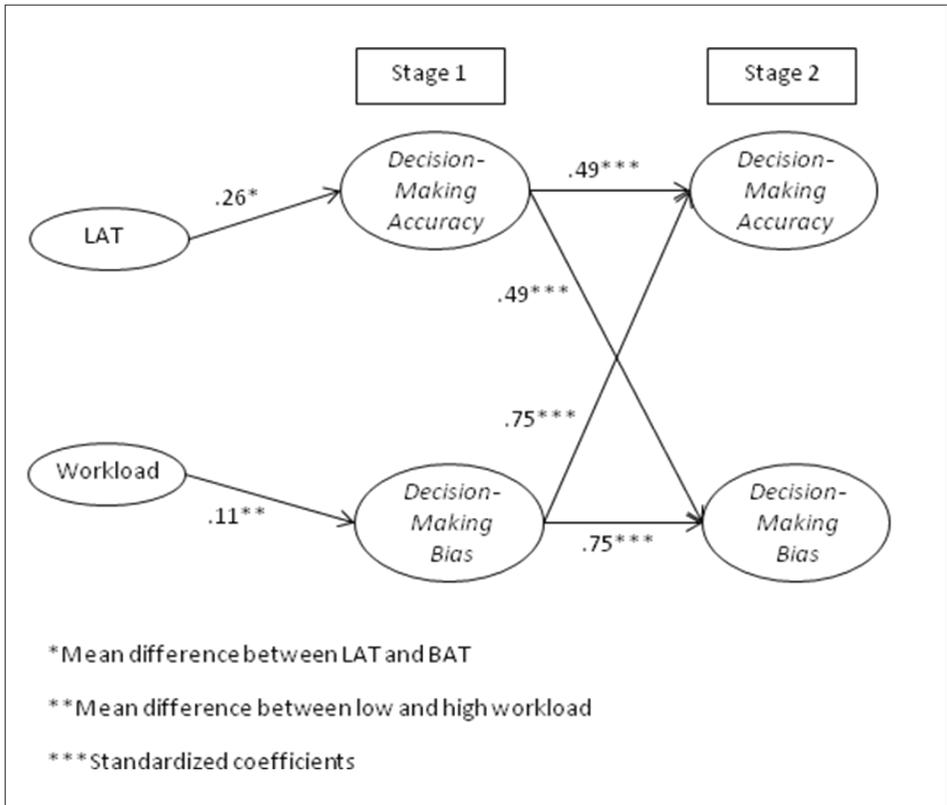
***Standardized coefficients

Figure 8. Schematic representation of Experiment 1 results.

Experiment 2

    Two 2 x 2 repeated-measures ANOVAs were used to analyze decision-making accuracy and bias during the first processing stage. System and workload were used as independent measures. Decision-making accuracy and bias were used as dependent measures. Results showed a significant main effect of system on decision-making accuracy, $F(1, 23) = 37.00$, $p < .001$, partial $\eta^2 = .62$. Participants' decision-making accuracy was significantly higher when they used the LAS ($M = .75$, $SE = .01$) rather than the BAS ($M = .58$, $SE = .01$).

A 2 x 2 repeated-measures ANCOVA was used to analyze decision-making accuracy and bias during the second processing stage. System and workload were used as independent variables. Decision-making accuracy and bias during the first processing stage were used as mediating variables. Decision-making accuracy and bias during the second stage were used as dependent measures. Results showed a significant mediating effect of decision-making bias during the first processing stage on decision-making accuracy, $F(1, 21) = 51.12$, $p < .001$, partial $\eta^2 = .71$, and bias, $F(1, 21) = 51.12$, $p < .001$, partial $\eta^2 = .71$ during the second processing stage. As participants' decision-making bias during the first processing stage increased, their decision-making accuracy during the second stage also increased ($r = .60$, $p < .001$), as well as their bias ($r = .60$, $p < .001$). Figure 9 shows a schematic representation of these results.
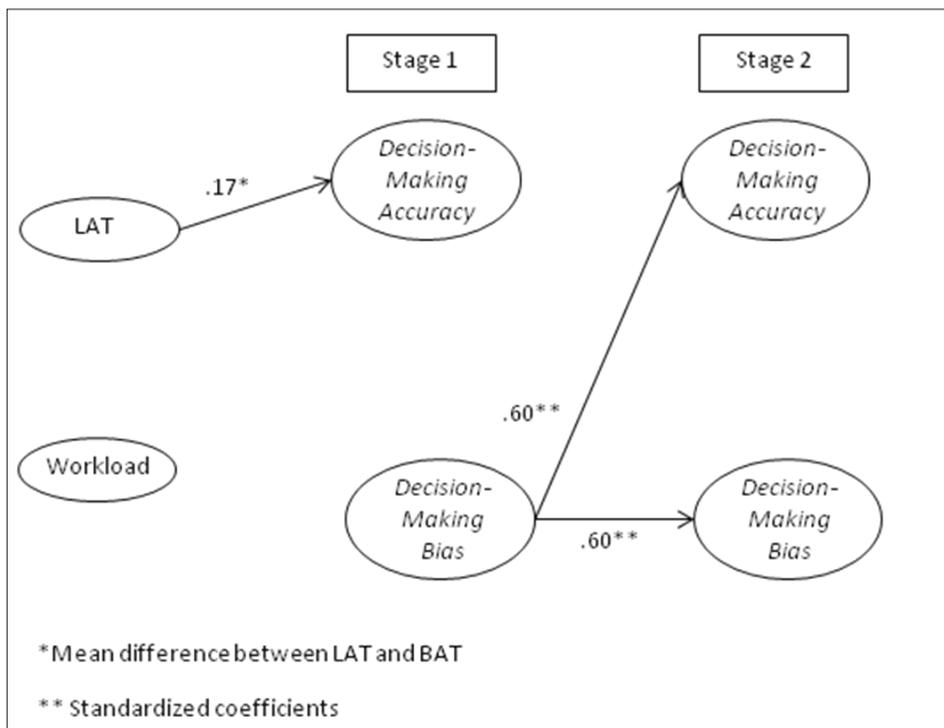

Figure 9. Schematic representation of Experiment 2 results.

Combined Results
  Two 2 x 2 x 2 mixed ANOVAs were used to analyze decision-making accuracy and bias during the first processing stage. TCI was manipulated between groups. System and workload were manipulated within groups. Decision-making accuracy and bias were used as dependent measures. Results indicated a significant inter-action effect between system and TCI on decision-making accuracy, $F(1, 52) = 6.27$, $p < .05$, partial $\eta^2 = .11$. Participant' decision-making accuracy improved with the use of the LAS and with the presence of TCI, but the latter had a significant effect only when participants interacted with the BAS. These results are shown in Figure 10.
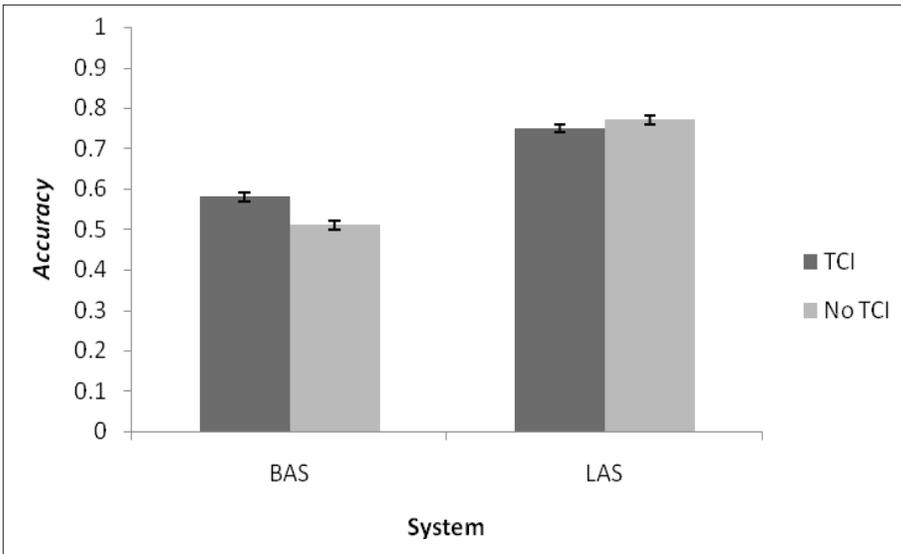
Figure 10. Decision-making accuracy as a function of system and task-critical information.

Results also showed a significant interaction effect between system and workload on decision-making accuracy, $F(1, 52) = 18.55$, $p < .001$, partial $\eta^2 = .26$. Participants' decision-making accuracy improved with the use of the LAS, particularly under high-workload conditions. These results are shown in Figure 11.
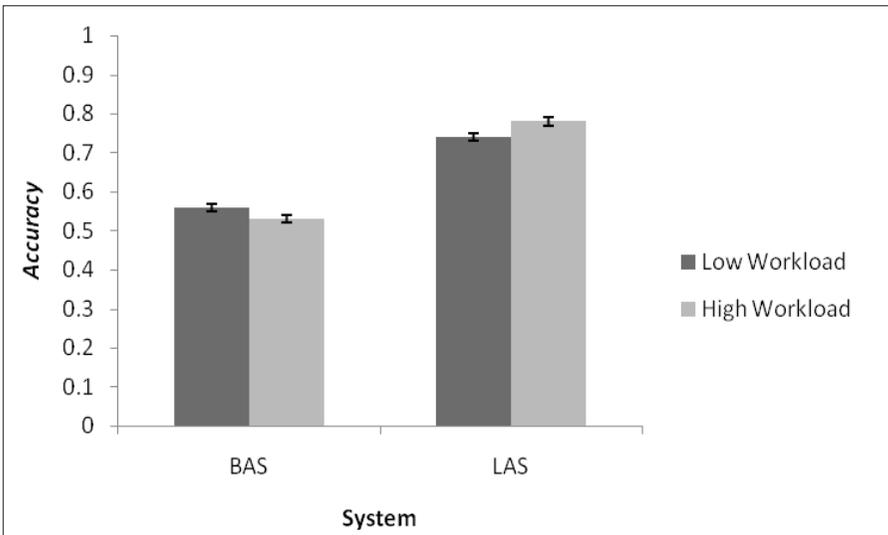

Figure 11. Decision-making accuracy as a function of system and workload.

Results indicated a significant interaction effect between TCI and workload on decision-making bias, $F(1, 52) = 17.94$, $p < .001$, partial $\eta^2 = .26$. The absence of TCI significantly increased participants' response bias during low-workload conditions. These results are shown in Figure 12.
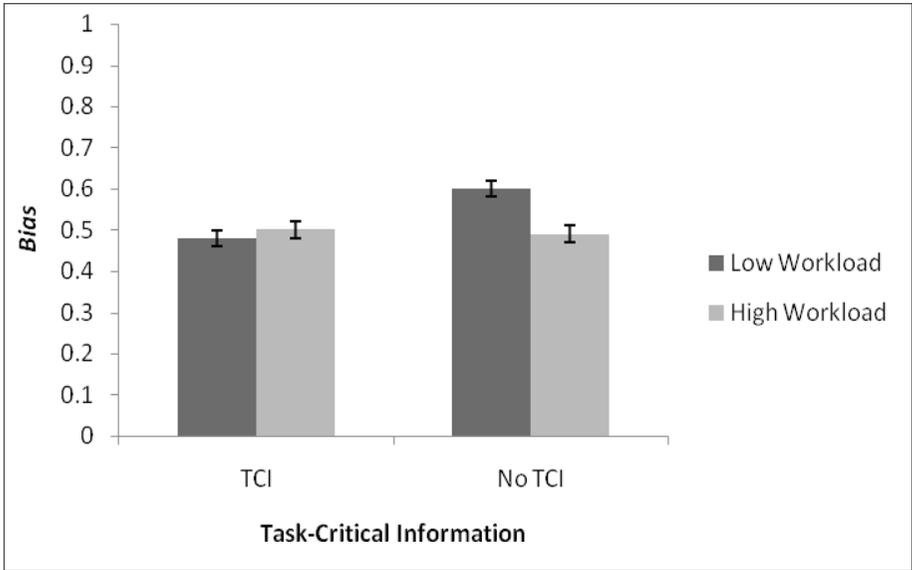

Figure 12. Decision-making bias as a function of workload and task-critical information.

A 2 x 2 x 2 mixed ANCOVA was used to analyze decision-making accuracy and bias during the second processing stage. TCI was manipulated between groups. System and workload were manipulated within groups. Decision-making accuracy and bias during the first processing stage were used as mediating variables. Decision-making accuracy and bias during the second stage were used as dependent measures. Results showed a significant mediating effect of decision-making accuracy during the first processing stage on decision-making accuracy, $F(1, 50) = 5.85$, $p < .05$, partial $\eta^2 = .10$, and bias, $F(1, 50) = 5.85$, $p < .05$, partial $\eta^2 = .10$ during the second processing stage. As participants' decision-making accuracy during the first processing stage increased, their decision-making accuracy during the second processing stage also increased ($r = .49$, $p < .001$), as did their bias ($r = .49$, $p < .001$). Last, results showed a significant mediating effect of decision-making bias during the first stage on decision-making accuracy, $F(1, 50) = 73.12$, $p < .001$, partial $\eta^2 = .59$, and bias, $F(1, 50) = 73.12$, $p < .001$, partial $\eta^2 = .59$ during the second processing stage. As participants' decision-making bias during the first processing stage increased, their decision-making accuracy during the second processing stage increased ($r = .68$, $p < .001$), as did their bias ($r = .68$, $p < .001$).

Discussion

Results were consistent with prior research and showed support for most of the hypotheses. As expected, workload and TCI had direct effects on response

bias during the first processing stage. Participants had a lower tendency to search for further information, regardless of whether alarms were true or false signals, when they had access to TCI, particularly under high workload conditions. As a result, workload and TCI had an indirect effect on response bias and decision-making accuracy during the second processing stage. As participants' bias during the first processing stage increased, their bias during the second processing stage also increased.

Workload also had a direct effect on decision-making accuracy during the second processing stage. Moreover, as predicted, TCI increased participant's decision-making accuracy during the first stage only when participants were interacting with the BAS. Last, as expected, LAT had a direct effect on participants' decision-making accuracy during the first stage. Participants had higher decision-making accuracy to alarms when they interacted with the LAS. This, in turn, increased their decision-making accuracy during the second processing stage. Participants made more accurate decisions while diagnosing system-status information regarding engine malfunctions when they interacted with the LAS.

Results showed support for using the proposed two-stage signal decision model. Using this model allowed researchers to examine the effects of workload, TCI, and LASs, on decision-making accuracy and bias during two separate processing stages: attention and decision-making. Results showed that workload and TCI had similar bias effects on both processing stages. However, results indicated that workload and TCI had differential indirect effects on decision-making accuracy during the decision making stage. The benefits of TCI during the attentional stage were counterbalanced by its indirect negative effect on the decision making stage by increasing response bias. The presence of TCI may have increased participants' self-confidence, thereby making them less likely to comply with the alarm system (Lee & Moray, 1994).

LAT, on the other hand, had only positive direct and indirect effects on decision-making accuracy, particularly during periods of high workload. One issue of concern with LASs is that they could bias decision-making, increasing pilots' errors of omission even when contradicted by reliable system-status information (Skitka, Mosier, & Burdick, 1999). However, the LAS did not have any effect on decision-making bias. Therefore, the findings of this research effort showed support for the superior advantage of using LASs to mitigate the cry-wolf effect, particularly under high workload conditions, above and beyond TCI.

These findings have potential practical applications regarding the design of integrated aviation displays and decision-support tools. From a policy standpoint, LAT could be implemented as the underlying algorithm upon which to provide pilots with advisories regarding conditions that require their attention, decision-making, and preventive or evasive actions. There are at least two methods to integrate LAT with pre-existing and future integrated aviation displays.

One approach is to gather relevant information about a particular potential problem from different sources and integrate it into a single decision variable

(Sorkin & Woods, 1985). This could be achieved using both linear and non-linear models for integrating data from different sources. The main purpose of this approach is to determine the probability of a potential problem given its magnitude as indicated by the integrated decision variable. An example of this approach that is particularly relevant to the EICAS system could be to integrate information from different engine parameters, such as pressure and temperature levels, into a single decision variable. Once the magnitude of this decision variable exceeds a low threshold, EICAS could emit a low-likelihood warning. If, on the other hand, the magnitude of this decision variable exceeds a higher threshold, EICAS could emit a high-likelihood alarm.

Another approach is to use Monte Carlo simulation, which is particularly applicable to future cockpit displays of traffic information. Yang and Kuchar (1997) developed a detecting algorithm that incorporates aircrafts' current state information, and utilizes future predictors, such as heading, speed, climb or descend trajectory, and intent information obtained from GPS and datalink communication. The detection algorithm follows a probabilistic model, which is estimated by conducting 500 Monte Carlo simulations per second. Each simulation introduces uncertainties in the estimation of the ownship's and surrounding aircrafts' current speed, altitude, and heading parameters. Furthermore, the simulation also introduces uncertainties in the ownship's and surrounding aircrafts' projected future trajectories. Based on these predictions, the algorithm counts the number of times a projected trajectory enters the ownship's protected zone, defined as a 10 nm in diameter and 2000 ft in altitude solid around the ownship. The algorithm then estimates the likelihood of a conflict by dividing the number of times a potential intruder enters the ownship's protected zone by the number of iterations (i.e., 500).

It is important to point out the potential limitations of this work and provide suggestions for future research. One of the limitations of this work was that the alarm system used in this research was a false-alarm prone system. Future research is needed to examine the potential benefits of LAT implemented in miss-prone systems as well (Wickens & Dixon, 2007). Another potential limitation of this research relates to the low level of fidelity of the simulated environment. Responding to alarms has implicit elements of risk and safety, which are difficult to achieve in a laboratory setting with low-fidelity simulation. Future research should focus on increasing the ecological validity of the simulated environment to better incorporate the implicit elements of risk and safety associated with dealing with potentially dangerous conditions. Also, an obvious limitation of this research was the nature of the sample. Previous research on decision-making suggests that people may make qualitatively different decisions based on their level of expertise (Klein, 1998). Further empirical research is needed to examine the generalizability of these findings using a sample of experienced certified pilots.

References

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology, 13*(3), 249-268.

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics, 43*(9), 1283-1300*.*

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300-2312.

Bliss, J. P., Jeans, S. M., & Prioux, H. J. (1996). Dual-task performance as a function of individual alarm validity and alarm system reliability information. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 1237-1241). Santa Monica, CA: Human Factors and Ergonomics Society.

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarms.* Hillsdale, NJ: Lawrence Erlbaum.

Bustamante, E. A. (2007). Using likelihood alert technology in cockpit displays of traffic information to support free flight. In *Proceedings of the 14th International Symposium on Aviation Psychology* (pp. 100-102). Dayton, OH.

Bustamante, E. A. (2008). The *a-b* signal detection theory model. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *68*(9-B), 6371.

Bustamante, E. A., Anderson, B. L., & Bliss, J. P. (2004). Effects of varying the threshold of alarm systems and task complexity on human performance and perceived workload. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (1948-1952). Santa Monica, CA: Human Factors and Ergonomics Society.

Comstock, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human pilot workload and strategic behavior research* (NASA Technical Memorandum No. 104174). Hampton, VA: National Aeronautics and Space Administration, Langley Research Center.

Edworthy, J., & Stanton, N. (1995). A user-centered approach to the design and evaluation of auditory warning signals: 1. Methodology. *Ergonomics, 38*(11), 2262-2280.

Elvers, G. C. (1997). The effects of correlation and response bias in alerted monitor displays. *Human Factors, 39*(4), 570-580.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System pilot response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*, 19-33.

Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Lee. J. D., & Moray, N. (1994). Trust, self confidence, and pilots' adaptation to automation. *International Journal of Human Computer Studies, 40*, 153-184.

Lehto, M. R., & Papastavrou, J. D. (1998). A signal-detection theory based perspective on design of warning. *Perceptual and Motor Skills, 86*, 720-722.

Maltz, M., & Meyer, J. (2001). Use of warnings in an attentionally demanding detection task. *Human Factors, 43*(2), 217-226.

Masalonis, A. J., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics, 46*(11), 1045-1074.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43*(4), 563-572.

Moray, N. (1981). The role of attention in the detection of errors and the diagnosis of failures in human-machine systems. In J. Rasmussen and W. B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 185-198). New York, NY:

O'Brien, J.V. & Wickens, C.D. (1997). Free flight cockpit displays of traffic and weather: Effects of dimensionality and data base integration. *In Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA: Human Factors and Ergonomics Society.

Papadopoulos, Y., & McDermid, J. (2001). Automated safety monitoring: A review and classification of methods. *International Journal of Condition Monitoring and Diagnostic Engineering Management, 4*(4), 1-32.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics, 40*(3), 390-399.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Pollack, I. & Madans, A. B. (1964). On the performance of a combination of detectors. *Human Factors,* 523-531.

Robinson, D. E. & Sorkin, R. D. (1985). A contingent criterion model of computer assisted detection. In R. E. Eberts & C. G. Eberts (Eds.), *Trends in ergonomics/human factors II* (pp. 75-82). Amsterdam, North-Holland: Elsevier.

Skitka, L.J., Mosier, K.L., and Burdick, M. (1999): Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991-1006

Sorkin, R. D. (1988). Why are pilots turning off our alarms? *Journal of the Acoustical Society of America, 84,* 1107-1108.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445-459.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human Computer Interaction, 1*, 49-75.

Stanton, N. A., & Baber, C. (1995). Alarm-initiated activities: An analysis of alarm handling by pilots using text-based alarm systems in supervisory control.

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist, 47*(4), 522-532.

Tumer, I. Y., & Bajwa, A. (1999). A survey of aircraft engine health monitoring systems. In *Proceedings of the 35th Joint Propulsion Conference, Aerospace Systems Condition Monitoring Session*, (pp. 1-7). Los Angeles, CA: American Institute of Aeronautics and Astronautics.

Wickens, C. D. (1987). Information processing, decision-making, and cognition. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 72-107). Oxford, England: John Wiley & Sons.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37*(3), 473.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science. 8*(3), 201-212.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors, 30*(5), 599-616.

Wogalter, M. S., Young, S. L., Brelsfor, J. W., & Barlow, T. (1999). The relative contributions of injury severity likelihood information on hazard-risk judgments and warning compliance. *Journal of Safety Research, 30*(3), 151-162.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics, 38*(11), 2371-2393.

Xiao, Y., & Seagull, F.J. (1999). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. *Proceedings of the Human Factors and Ergonomic Society 43rd Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 256-260.

Yang, L. C., & Kuchar, J., K. (1997). Prototype conflict alerting system for free flight. *Journal of Guidance, Control, and Dynamics, 20*(4).

Locus of Control and Self-Attribution as Mediators of Hazardous Attitudes among Aviators: A Review and Suggested Applications

John E. Stewart
U.S. Army Research Institute
Fort Rucker Research Element, Building 5100
Fort Rucker, AL 36362-5354
Email: john.e.stewart@us.army.mil
Telephone: 334.255.9109

Abstract

Locus of control (LOC) predicts many attitudes and behaviors consistent with safety and risk taking. Among these are perceptions of risk, attributions of skill vs. luck, self-discipline, and information seeking in hazardous situations. Over the past two decades, researchers have examined the relationship of LOC to hazardous attitudes, involvement in hazardous events, and other variables related to aviation safety. Research has found aviators to be significantly more internal in LOC than the general population, internality being associated with self-attributions of competency and self-confidence. This research has been cross-sectional and has typically employed small samples of civil aviators. Questions have been raised about psychometric properties of earlier (ipsative) versions of the hazardous attitude scales, and the resulting dependency between the five attitudinal categories. This review examines extant LOC research in aviation psychology, most of which addresses LOC in the context of hazardous attitudes. It examines concepts from attribution theory, (e.g., optimism bias; illusion of unique invulnerability), and argues that these are consistent with the processes underlying the maintenance of LOC and hazardous attitudes. It is recommended that integration of LOC and attribution theory should provide an enhanced explanation of the motivational bases for risk taking and decision making among aviators.

Locus of Control and Self-Attribution as Mediators of Hazardous Attitudes among Aviators: A Review and Suggested Applications

"Don't be a show-off. Never be too proud to turn back. There are old pilots and bold pilots, but no old, bold pilots."   E. Hamilton Lee,[1] 1949

---

1. E. Hamilton "Ham" Lee began his long and distinguished career as an instructor pilot during World War I. After leaving the Army Air Corps, he flew the airmail for United Air Services, later United Airlines. The "old pilots, bold pilots" statement was made on his retirement from United Air Lines. Ham Lee did indeed become an old pilot. On his 100th birthday, in 1992, he piloted a restored United Airlines DC-3 from the left seat. He died in 1994 at the age of 102.

Overview

Locus of control and risk perception. Locus of control of reinforcement (LOC) consists of a set of expectancies that outcomes are influenced by one's own efforts (internal) or by environmental forces beyond one's control (external). LOC has been applied to a host of settings, including traditional classroom learning situations, industrial safety, medicine, and, more recently, aviation. The application of LOC research to aviation has been limited, and consequently we know little about variation in LOC among student, journeyman, or senior pilots. Furthermore, little is known about how military and civilian pilots may differ on this dimension. Expanding this knowledge base could provide additional insight as to changes in LOC as a result of lessons learned from hazardous experiences during one's lifetime. Related constructs which could be employed in aviation psychology research include psychometrically-refined versions of scales purporting to measure hazardous attitudes and involvement in hazardous incidents (Hunter, 2005), as well as constructs from attribution theory relating to the perception of one's own risk orientation relative to similar others (e.g., optimism bias). One premise of this paper is that a case can be made that the expectancy theory foundations of LOC are closely linked to attribution theory, with the main difference being one of perspective (actor/observer). This common theoretical foundation should provide an integrated attribution-based model for understanding decision-making and risk taking among aviators.

Purpose of this review. The goal of the present review and discussion is to illustrate that understanding human performance situations in aviation can be enhanced by applying well-established constructs from social psychology, namely, LOC and the self-attribution of ability. Many of the skills required for flying an aircraft depend upon the development and sustainment of a sense of control, as well as the ability to anticipate situations where one's sense of control and self-attributions of ability may be challenged by external circumstances. Similarly, the tendency to endorse an internal LOC as the desirable state in a person may not hold true across all situations. The five hazardous attitudes (Berlin, et al., 1982; Federal Aviation Administration, 1991) refer to specific attributional biases that can result in poor decisions, which can have potentially fatal consequences. These "five deadly sins" suggest that pilots can get into trouble if they develop unrealistic perceptions of their abilities, the environment, and control over their fate (Stewart, 2006). These attitudes seem to be closely related to self-attributions of ability. They are: *antiauthority* (these rules don't apply to me); *macho* (I can do it!), *impulsivity* (I don't have time for this); *invulnerability* (this can't happen to me), and *resignation* (what's the use ?). The review will pinpoint limitations of past research (e.g., small samples that limit generalizability, psychometrically flawed scales), which call for further research, and will recommend future approaches to the study of LOC and attribution theory in aviation settings.

The majority of participants studied were general aviation (GA) pilots. At the present time, we do not know the effects of adding the risks associated with combat to the inherent risks of civil aviation settings on a pilot's sense of personal control. To date, only one published study (Joseph & Ganesh, 2006), compared military and civilian aviators in terms of LOC, In brief, a lot remains to be learned about the effects of LOC and related constructs on aviation decision making (ADM).

Locus of Control of Reinforcement

Rotter LOC scale. Research over the past three decades has shown Rotter's (1966) LOC construct to be predictive of a broad range of human behaviors. LOC had its origin in Rotter's (1954) social learning theory, which is an outgrowth of reinforcement and expectancy theories. Rotter's social learning orientation conceptualizes personality as more of a situational *state* than a dispositional *trait*, comprising self-attributions and expectancies, which can change over time. LOC is *not* a personality trait. Expectancies are influenced by situational factors and are therefore changeable, whereas stable dispositions tend to be resistant to change across situations. A more extensive discussion of the state vs. trait controversy as it applies to LOC can be found in Stewart (2006).

Internality of LOC implies that the person believes that outcomes are due to *skill,* whereas externality connotes a belief that these are due to *chance*. When one mentions LOC, the Rotter (1966) scale comes to mind. This scale consists of 23 items, (plus 6 filler) using a forced-choice format with only external responses being scored. Scores can range from 0 to 23.

Aviation Safety LOC scale (ASLOC). A large number of LOC scales, many of them adapted to specific age groups (e.g., children) and settings (e.g., health, safety, traffic) have evolved since the original Rotter scale. Hunter (2002) has developed and validated the ASLOC scale, specific to aviation, predicated upon the finding that the predictive validity of LOC against external criterion measures is enhanced when the scale items correspond to specific settings and target behaviors (Montag & Comrey, 1987).

The ASLOC was derived from an industrial safety LOC scale (Jones & Wuebker, 1985). The ASLOC combined index ranges from 20 to 100, (20 items; 5-point scale) with higher scores keyed to internality. Based upon the results of a principal components factor analysis, which failed to reveal a single internal-external continuum, Hunter (2002) broke the scale down into internal and external subscales, each consisting of 10 items. Two theoretical approaches for scoring LOC exist. One approach depicts LOC as two opposite poles of a single unitary dimension; another argues that internality and externality are separate constructs. Hunter contends that the bulk of the literature and his own data support the latter position. The reader is referred to Collins (1974) and Montag and Comrey (1987) who review and discuss in detail the issue of dimensionality of LOC.

Do specific experiences affect LOC? An important research issue is the degree to which LOC is modifiable due to direct intervention or life experiences. Considering that LOC is based on expectancy theory, and experiences affect subjective expectancies, this would seem a reasonable assumption. Research involving training interventions designed to increase one's sense of personal control (Duke, Johnson, & Nowicki, 1977; Lynch, Ogg, & Christensen, 1975) have demonstrated some degree of success. These findings imply that LOC is a state variable that can be influenced by the situation.

Cross-sectional studies on high school and college students (Phares, 1976; Twenge, Zhang, & Im, 2004) paint a less optimistic picture. These studies have shown an annual trend toward increasing externality of LOC with successive age cohorts over time. Twenge, Zhang and Im found that the average college student in 2002-2003 was more external on the Rotter LOC scale (M = 11.38) than 80% of his or her counterparts from the 1960s (M = 8.70). From an evaluative standpoint, implications of these findings are negative, as externality is associated with alienation, self-attributions of powerlessness, and lower achievement. Hunter (2002), in a validation study, using a cross-sectional sample of 477 civilian pilots, found a statistically significant trend of increasing internality with age, though not with flight experience. We should note that these were not longitudinal comparisons and that many historical, cultural, and experiential differences between cohorts can contribute to these trends.

LOC and attribution theory: two convergent paradigms. Although the linkage between LOC and attribution theory may not be obvious at the outset, a brief overview of the attributional notion of locus of cause (Heider, 1958) reveals a high degree of conceptual similarity. Both are concerned with the perception of personal causation (locus of cause). The difference is whether the behavior is seen though the eyes of an outside observer, or whether the actor is observing the effects of his or her own behavior. A central assumption of Heider's attribution theory is a need to perceive human acts as originating inside the person (dispositional) or from influences outside the person's control (situational). According to Heider, two influences contribute to how one perceives the causality of another's behavior: *volition*, or how hard the actor tries (assuming appropriate ability), and *situation*, or the degree to which the actor seems compelled to engage in the behavior. Thus, the locus of cause will appear *internal* if the actor seen as having the *ability* to perform a task also puts forth the *effort* to accomplish it. Contrariwise, locus of cause should appear *external* if the actor is seen as *not* having the ability to accomplish the task. If the actor who has the ability fails at the task, this outcome is assumed to be due to lack of trying. If the actor did not have the ability, failure is attributed to external circumstances, as the task was too difficult. This is analogous to internal and external LOC, since the actors make the same inferences about the consequences of their own behavior.

A pioneering experiment (Jones & deCharms, 1958) illustrates the relationship between perceived locus of cause and LOC. The perceiver (i.e., participant in the experiment) was deprived of a reward on a two-person problem-solving task due to the performance of the stimulus person (actually a confederate of the experimenter). The degree to which the person's failure was due to lack of effort (internality) or being confronted with an impossible task (externality) was the independent variable. The stimulus person was evaluated more negatively when his failure was due to (perceived) lack of effort than when it was due to the difficulty of the task before him. From the standpoint of LOC, the stimulus person either had the ability to control the outcome of the task, but failed to exercise control, or did not have the ability, and thus could not exercise control. LOC, in short, can be conceptualized simply as a form of self-attribution of the causation and consequences of one's own behavior.

Relationship of LOC to Hazardous Attitudes and Attribution

   Early research using ipsative HAS and Rotter LOC scales. The question emerges as to whether LOC is closely correlated with all or part of this typology. One hazardous attitude, i.e., resignation, was originally labeled *external control* by Jensen and Benel (1977) during the original identification of the five hazardous thought patterns. The scale purporting to measure these attitudes, the Hazardous Attitudes Scale (HAS), has recently come under criticism because it employs an ipsative, forced-choice format (Hunter 2005). This format creates dependency between the five factors, in that endorsement of items determines lower scores on the other four. Lester and Bombaci (1984) examined the relationship of LOC to the five hazardous attitudes, as indicated on the older, ipsative HAS, using a sample of 35 male GA pilots. This study found that invulnerability was the predominant hazardous attitude (43%) followed by impulsivity (20%) and macho (14%). No participants fell into the remaining two categories (Percentages did not sum to 100% because some participants did not fall into any of the five categories). Macho aviators were the most internal of all on the Rotter LOC scale (M = 3.4; n = 5), followed by invulnerability (M = 8.1; n = 15) and impulsivity (M = 10.3, n = 7). Compared to large-sample archival data on LOC none of these means could be called external; only the impulsivity group scored close to the median (11.5) on the Rotter LOC scale. Comparisons with a benchmark or norm are difficult because LOC scores appear to be susceptible to situational and historical influences (Phares 1976). The extreme internality of the macho group is intriguing, though this subsample is small. Nevertheless, mean scores for macho and invulnerable, compared with those of the impulsive participants, seem consistent with much of what is known about LOC. However, impulsivity as indicated on the HAS might not be the same as that measured by personality inventories. The authors found no significant relationship between the impulsivity dimension from the HAS and the impulsivity scale of the 16 Personality Factor Questionnaire (16 PF [Cattell, Eber, & Tatsuoka, 1970]), but found that macho participants scored significantly higher than invulnerables on the integration/self concept control scale of the 16 PF.

   The results of Lester and Connolly (1987) parallel closely those of Lester and Bombaci (1984). Again the predominant hazardous attitude, among a subsample of 89 participants, was invulnerability (39%) followed by impulsivity (24%) and macho (19%). Approximately 13% of participants displayed no dominant hazardous attitude. As in Lester and Bombaci (1984), resignation and antiauthority did not emerge. The investigators also found that the three attitudinal dimensions were significantly and positively correlated, possibly an artifact of the ipsative format. Lester and Connolly contrasted the LOC scores of macho (M = 7.1), invulnerable (M= 9.8) and impulsive (M = 10.1) participants. Macho participants were significantly more internal than invulnerables ($p < .05$), and more likely to report involvement in hazardous situations or accidents ($p < .05$), though showing more conscientiousness on the 16PF ($p < .05$). The authors also administered a version of the same HAS, which employed a 6-point scale rather than a forced-choice format, to another subsample (n = 60). The participant was asked to rate the degree to which the irrational behaviors described in the questionnaire were

typical or atypical of their own behavior. This was considered an index of propensity toward irrational decision-making. The investigators found that pilots with self-reported better judgment were significantly more internal (*M* = 7.2) than were those indicating poorer judgment (*M* = 9.8; *p* <.05). Again, neither score can be classified as external, which reaffirms internality as the dominant LOC orientation among pilots. Finally, there were no differences between the two groups as to self-reported involvement in aviation accidents. Of this subsample, 33.6% reported involvement in a near-accident, 11.4% in an actual accident or incident, and 41.4% in some dangerous aviation incident.

LOC and the self-serving bias. Wichman and Ball (1983) examined LOC in the context of self-attribution, rather than ADM style as reflected in the HAS. These investigators surveyed 334 GA pilots (three subsamples surveyed at a flight instructors' clinic, fixed base operators and airports) concerning safety attitudes and practices, using the Rotter LOC scale. This study is important because it linked LOC to the self-serving bias (Ross, 1977), a special case of attribution theory in which the person, or actor, attributes positive outcomes to internal factors (effort), and negative outcomes to external factors (luck). Consistent with other research, they found pilots to be significantly more internal than the general non-aviator population in the United States. Comparison between Rotter's (1966) sample (*M* = 8.3) and three subsamples (i.e., academic classes) of GA pilots surveyed by the authors, showed the student pilots in the three classes to be significantly more internal in LOC (*M*s = 6.9; 6.1; 6.2, respectively). In addition to the Rotter scale, pilots were asked to respond to four questions (9-point scale) concerning their self-perceptions of flight safety, skill, and chances of having an accident, compared to "aviators with the same number of flying hours and exposure as you have." The fourth question was concerned with their perceptions of flight safety in general. The pilots showed evidence of a self-serving attributional bias toward their own skill and safety, believing themselves to be significantly more skilled and less likely to have an accident than other pilots of similar experience. Pilots with more flight hours evidenced more internality of LOC and attributional bias than their less experienced counterparts. However, these experienced pilots, though tending toward greater internality and self-serving bias, were *not* more cavalier than less experienced pilots on matters of safety. On the contrary, they were more likely to attend an FAA safety clinic. This behavioral measure implied that they not only believed themselves to be safer than the average pilot, but also behaved in a way consistent with these beliefs. This suggests that internality and high self-confidence (as reflected in their self-perception of having better than average piloting skills) can indeed promote behaviors associated with safe flying.

Hunter's revisions of the HAS scales and development of the ASLOC. Hunter (2005) based the derivation of the non-ipsative version of the HAS (New-HAS) on research by Holt, et al. (1991) who designed a Likert-based instrument for automobile drivers. For the validation study Holt, et al. used an equivalent form of the New-HAS with content appropriate for drivers (for the sake of convenience; pilots were scarce among student participants). The scale was administered to a sample of 238 undergraduates. Holt, et al. found four factors corresponding to the hazardous thought patterns: macho, impulsivity, antiauthority, and resignation (external LOC). The invulnerability factor was represented by the opposite pole of the dimension: low worry/anxiety about one's driving. A sixth factor seemed to reflect per-

sonal feelings of competence and self-confidence about one's driving. Factor scores correlated significantly with several criteria, such as accident involvement, drinking while driving, moving violations, and seat belt use. Impulsive (impatient) and antiauthority drivers reported significantly more problems with increased insurance rates, drinking while driving, moving violations, decreased seat belt usage, as well as accidents and incidents, than did those in the other categories. Invulnerable (i.e., low worry/anxiety) drivers reported more parking tickets than others, reduced seat belt use, and negative reactions from other drivers. Resigned drivers reported problems with not using seat belts and receiving negative reactions from passengers. By contrast, the confident drivers, though they did drink and drive, reported fewer accidents than the others, and positive reactions from passengers regarding their driving.

Using the New-HAS, Hunter (2005) conducted a web-hosted survey of GA and commercial pilots to determine the factor structure of the scale. Exploratory factor analysis was followed by a Varimax rotation to simple structure. Analysis revealed six factors, which paralleled those found by Holt, et al. These were macho, resignation, antiauthority, worry/anxiety, impulsivity, and self-confidence. Hunter discovered that items of similar content on the Old-HAS (i.e., ipsative HAS) and New-HAS yielded low correlations, probably due to psychometric differences between the two scales. Hunter (personal communication, July 10, 2008), states that invulnerability, in the form of low worry/anxiety, is still one of the predominant hazardous attitudes, and that, indeed, the hazardous thought patterns seem to have survived the rigors of factor analysis.

Hunter (2002), in his efforts to determine construct validity of the ASLOC, found a significant, negative correlation between the internal subscale and self-reported involvement in hazardous events, ($p < .01$) as indicated by the Hazardous Events Scale (HES; Hunter, 1995). He defined a hazardous event as an accident or an incident, which could easily become an accident (e.g., running low on fuel). One must note that the HES is a self-report measure. Other safety-related LOC research studies(e.g., Arthur, Barrett, & Alexander, 1991) indicate that self-reports are much less sensitive criteria than archival records of accidents. Hunter (2006), in another web-based study, asked 369 GA pilots to estimate the risk involved in a set of hypothetical flying scenarios, for fictitious other pilots and for themselves. He found that responses to the HES were significantly and negatively correlated ($p < .05$) with accuracy of the perceived risk of flying. Perceived risk accuracy was defined as the degree to which the respondent believed that GA flying was safer than driving under comparable conditions (it actually is not). This was considered a measure of inaccurate perception of flight safety. Scores on the ASLOC Internal subscale were correlated positively and significantly ($p < .05$) with the perception of high flight risk. The study complements Hunter's other efforts, in that it demonstrates some degree of construct validity to the HES and shows that internality on the ASLOC is associated with the accurate perception of the scenarios as involving high risk. Hunter's finding is of potential interest to aviation safety; it indicates that internal LOC may be a predictor of vigilance and avoidance of unreasonable risks.

Joseph and Ganesh (2006) administered Hunter's ASLOC scale to 101 Indian aviators, comprising 50 military (Indian Air Force) and 51 civil pilots. No significant correlations were found between demographic variables (e.g., age and flying hours) and scores on the internal and external subscales, or on the combined ASLOC scale. Civil (i.e., transport) pilots were found to have significantly higher scores ($p < .02$) on the internal subscale of the ASLOC ($M = 36.63$, $SD = 4.97$) than military pilots ($M = 34.44$, $SD = 5.00$). Among the military aviators, fighter pilots ($M = 35.26$, $SD = 3.95$) were significantly more internal ($p < .05$) than were helicopter pilots ($M = 32.60$, $SD = 4.82$). The investigators postulate that the differences in LOC between civil transport, military fighter, and military helicopter pilots could reflect different adaptations to their operational environments. Civil transport operations are more routine than military operations and usually go as planned. By contrast, fighter operations are less routine and less predictable than transport, and helicopter operations are the least predictable, often responding to mission taskings on short notice. One problem with interpreting the results of this research is due to the use of multiple *t*-tests, in which some comparisons do not appear to be independent. For example, comparing civil to military pilots and transport to fighter pilots are not independent comparisons, since all but six transport pilots were civil aviators, and fighter pilots were a subsample of the larger sample of military pilots (fighter + helicopter). This study is important because it compares LOC scores among civil and military aviators, and secondarily because it represents a cross-cultural administration of the ASLOC.

Hunter (2008) reported preliminary results from a sample of 280 U.S. Army senior aviators, all rotary wing pilots, who were enrolled in a training course for Aviation Safety Officers. He then compared their ASLOC scores to those of a sample of 477 civil pilots (see Hunter, 2002). Hunter found that scores on the internal subscale were significantly higher ($p < .03$) for civil ($M = 38.80$, $SD = 4.34$) vs. Army ($M = 37.99$, $SD = 5.94$) aviators. On the external subscale, Army aviators ($M = 21.42$, $SD = 6.38$) scored significantly higher ($p < .01$) than did civil aviators ($M = 17.20$, $SD = 3.79$). Thus, on both subscales, Army aviators were significantly more external in LOC than civilian aviators. The underlying reasons for these differences are unclear, since the two independent samples differed in many ways. Nevertheless, it is interesting to note a parallel between these findings and those of Joseph and Ganesh (2006), in that civil pilots showed more internality on the ASLOC than did military pilots. Generalization from these preliminary findings should be made with caution, since the vast majority of Army aviators are rotary wing rated, whereas most GA pilots are fixed wing rated.

LOC, Confidence, and Sense of Invulnerability

Overconfidence and the Illusion of Invulnerability

When is internality a bad thing? Unrealistically high expectations of personal control may be maladaptive. One recent study (Anderson, Hattie, & Hamilton, 2005) purports to show evidence that among school children, moderate levels of LOC are more adaptive than are extremely high or low levels. Rotter (1966) acknowledged that those holding extreme internal beliefs might be as maladjusted as those holding extreme external beliefs. There is no empirical evidence to support any relationship between extreme internality and overconfidence in pilots; that is to say, what data exist, do not show a relationship of increasing invulnerability

being associated with increasing internality. Instead, it appears that those categorized as invulnerable on the Old-HAS manifest moderate rather than extreme internality scores (Lester & Bombaci, 1984; Lester & Connolly, 1987). Recall also that Hunter (2002; 2006) found internality to be associated with fewer self-reports of involvement in hazardous events as well as with accurate perception of risks. His factor analysis of the New-HAS scale (2005) did not reveal invulnerability as a unitary hazardous attitude. Instead, two factors emerged that could be characterized as self-confidence and worry-anxiety, with invulnerability coordinated to high self-confidence and low worry-anxiety.

Can confidence become invulnerability? Dunning, Heath, and Suls (2004), in a comprehensive monograph, demonstrated the pervasiveness of overconfidence and the illusion of control across institutional and occupational settings. Some evidence indicates that perceived control is a key mediator of one's optimistic bias (Helweg-Larsen & Shepperd, 2001; Klein & Helweg-Larsen, 2002). Klein and Helwig-Larsen performed a meta-analysis of 21 studies, which had explored the relationship between perceived personal control and the optimistic bias. Perceived control was found to have an impact upon risk perception. Those who believed that they could prevent the occurrence of negative events believed themselves to be at less risk than most others. What we do not know from these results is whether some of those who believed themselves at less risk than others behaved in a manner consistent with this perception (e.g., actively took preventive measures). Wichman and Ball (1983), it should be recalled, did find some behavioral differences between pilots differing in LOC and confidence.

Though evidence is accumulating that sense of personal control and overconfidence are major self-serving attributional biases in Western society, there is scant evidence as to how LOC mediates these biases. On its face, the term *sense of personal control* seems similar to internality, but this is not to be assumed in the absence of empirical support. The meta-analysis by Klein and Helweg-Larsen (2002) is the most pertinent research in the context of what has been discussed thus far. The authors define optimistic bias as a perception that one's risks of negative outcomes (e.g. accidents, alcoholism, disease, divorce) are less than those of similar others. The research literature has shown this bias to be quite robust (Weinstein, 1987; Weinstein & Klein, 1996).

Longitudinal Trends in Sense of Personal Control
There is some evidence in the research literature (reviewed by Helweg-Larsen & Shepperd, 2001) that prior experience with a negative outcome can have a moderating effect on the optimistic bias. In short, persons can perceive themselves as being at greater risk than before, when it is discovered that bad things can happen, even if one is competent and takes precautions.

If this relationship actually exists, then it would be possible to design training programs to minimize the tendency for the perception of "being in control" in moderately experienced pilots from becoming the "illusion of invulnerability." According to Helweg-Larsen and Shepperd (2001), there has been no empirical research into why prior experience reduces the optimistic bias. However, they do see per-

sonal control as a potential moderator. The authors found evidence that a declining sense of personal control over negative events causes estimates of personal vulnerability to increase. This trend was found in the meta-analysis regardless of whether the sense of personal control was measured as an individual difference variable (e.g., LOC), self-ratings of controllability of a target event, or inferred from a person's prior experience. One problem pointed out by Helweg-Larsen and Shepperd, which makes these findings hard to explain, is the absence of any experimental research on the relationship between perceived control and prior experience on the optimistic bias.

Wilson and Fallshore (2001) surveyed 160 commercial and GA pilots, finding evidence for self-serving or optimistic biases. Participants rated themselves, compared to other pilots, as less likely to experience accidents due to inadvertent flight into Instrument Meteorological Conditions (IMC) and overestimated their own ability to avoid and to escape IMC. Age and flight hours were unrelated to participants' estimates of the likelihood of having an accident, but were significantly related to estimates of the ability to escape from inadvertent IMC. Flight hours predicted estimates of the ability to avoid IMC altogether. Wilson and Fallshore admit that the latter findings were surprising, in that one would expect pilots to become more circumspect about hazards such as icing and IMC with increasing experience. However, one could also argue that more experienced pilots are more likely to have dealt successfully with IMC hazards, simply because they have learned how to do so. The reader should also recall that Wichman and Ball (1983) found that optimistic perceptions of one's own abilities could be associated with positive attitudes and behaviors toward flight safety. Actual flight hours are not reported, making it difficult to determine just how experienced these pilots were.

Illusion of Unique Invulnerability
Perloff and Fetzer (1986) posit the illusion of unique invulnerability as a cognitive process mediating excessive risk taking in people who should know better. This is similar to the optimistic bias, but based more on cognitive consistency than on self-attribution. However, the assumptions and predictions that it makes seem to be highly similar. Perloff and Fetzer point to a mode of inconsistency resolution not unlike differentiation (Abelson, 1959). That is to say, actors make inappropriate and stereotyped interpersonal comparisons between themselves and those who fall victim to the hazards of the target behavior. Thus those who are seriously injured or killed in aviation accidents may be seen as different from those who survive the same hazards; they may be perceived as less intelligent, less skilled, or poorly trained, from the perspective of the perceiver. As a means of maintaining cognitions that this is a *just world*, victims of accident and injustice are derogated. The less the evidence that the victim was negligent or incompetent, the greater is the need for derogation. Consequently, the need to maintain cognitive consistency can compound the degree of injustice. This just world phenomenon is a well-established process known for decades among social psychologists (Lerner, 1997, 2003; Lerner & Miller; 1978).

The point of the preceding discussion is that Perloff and others explain the illusion of unique invulnerability based upon actors' needs to see themselves in control of the situation and not at the mercy of external forces. This in turn serves as an anxiety reducing mechanism. ('It can't happen to me; I am a much better pilot

than they are; it happened to them because they are much worse pilots than I am'). Unfortunately, the illusion of unique invulnerability may do more than just reduce anxiety and bolster one's self esteem; it may lead to excessive risk taking.

The concept of invulnerability seems closely linked to the expectancy of personal control. Evidence shows that people, or at least participants in experiments, systematically distort their degree of control over positive (successful) outcomes, giving them a heightened sense of personal control (Alloy, Abramson & Viscusi, 1981). This has been found to be particularly true of participants who experience repeated successes. For aviators the "downside" may be the point where a pilot begins to overestimate his or her abilities, which could be dangerous if the aviator is relatively inexperienced in actually dealing with potentially hazardous situations. This illusion is maintained by seeing similar others as less in control and underestimating the power of environmental factors (Wichman & Ball, 1983). Are less experienced pilots more susceptible to these cognitive biases than their more experienced counterparts? Is there a level of experience where this susceptibility is greatest? An answer to this question awaits future research efforts.

It would seem that LOC and the distortions that result in the illusion of control are parts of the same social learning process. A person strives to achieve a goal, learns that his or her efforts have been successful, and perceives a causal connection between effort and outcome. After repeated successes at the task, the expectancy of future success should increase. This conceptualization should provide a better view of the social learning-expectancy theory origins of LOC. One must reconcile this, however, with the small amount of research that indicates that pilots with high internality of LOC tend to be more accurate in their perception of risks than those of moderate internality; perhaps the issue in this case is the degree of congruency between one's sense of personal control and the actual level of risk in the environment.

Discussion

Limitations of the Research
   Range restriction and sample size. Besides recurrent sample size problems, another limitation of LOC research in aviation settings is the strong possibility of a restriction in range of LOC scores among pilots. The small amount of data thus far shows that internality is the prevailing LOC orientation among pilots. This has been shown for both the Rotter LOC and the aviation-specific ASLOC.

Research addressing LOC in the context of aviation has concentrated primarily upon its relation to the five hazardous attitudes. Initially, researchers attempted to validate the Rotter LOC scale against the older, ipsative HAS scale. These early studies suffered from small sample size, along with the conceptual and psychometric limitations of the older version of the HAS. Subsequent web-hosted research by Hunter (2002, 2005) developed and validated a new version of the LOC scale, with content items specific to aviation, as well as the Likert-based New-HAS and HES scales. The five hazardous attitudes fell out of the

factor analysis, along with a sixth, self-confidence. Only three (macho, invulnerability, impulsivity) were found by Lester and Bombaci (1984) and Lester and Connolly (1987). This disparity in findings could be an artifact of the ipsative nature of the scales used by the latter researchers.

Hunter (2005), one should recall, found self-confidence and worry/ anxiety as independent factors. This implies that an aviator can be highly self confident but at the same time manifest a moderate amount of worry or concern about the hazards that all pilots must face. On the other hand, one could postulate that a highly self-confident pilot who does not show any concern about these same hazards may have fallen victim to the illusion of invulnerability.

Military vs. civil aviation. Military pilots may differ in many ways from GA and commercial airline counterparts. These differences may be dispositional or situational. Besides concurrent validation of the ASLOC, New-HAS and HES scales on samples of military pilots, another advantage of research in military settings is the availability of archival accident reports, which could be content analyzed for wording corresponding to such variables as hazardous attitudes and risk orientation. Wetmore, Bos, and Lu (2007) have recently conducted such a case-based analysis for civil aviation accidents, using the five hazardous attitude categories as criteria. The analysis revealed that, similar to previous research using GA aviators, invulnerability was the most prevalent hazardous attitude associated with accidents (80% of the accident pilots). Wetmore, Bos, and Lu then extrapolated from the analysis to posit instructional strategies that would best deal proactively with hazardous attitudes among student pilots, which, they reasoned, should be evident to the flight instructor. In short, one's dominant hazardous attitude could be a diagnostic, for which a training strategy is a possible antidote. Although this study did not address psychometric issues in the use of the Old-HAS, it did suggest a much-needed proactive approach to accident analysis, which should be useful to military aviation.

Perhaps there is a time and place for potentially hazardous attitudes in wartime. Situations exist in which the "Can Do" attitudes prevalent in military aviation and other high-pressure professions may be appropriate and necessary; in others (e.g., peacetime aviation), much more cautious, deliberative attitudes may be appropriate. The key is flexibility. LOC and the hazardous attitudes should be variable and situation-appropriate. Making situation-specificity salient to pilots at all stages of their careers should become an important goal of aviation training. Risks acceptable in a military combat situation, where aircraft and crews may be expendable, are obviously not acceptable in non-combat or peacetime situations.

Critical Needs for Future Research
The need for external criteria. The use of self-report measures, usually completed by the same respondents, imposes limitations on new knowledge. Now that new, psychometrically enhanced HAS scales have been developed, the next step is to validate them against independent, behavioral criteria. Several approaches suggest themselves, including accident scenario training in the simulator, in which specific judgmental errors can be recorded by observers blind to the participants' LOC, HAS, or other self-report scores. Do macho pilots show distinctive error patterns compared to those who are anti authority or high in self-confidence? Do

internals (in LOC) process hazard-related data more effectively than externals? How do these different orientations combine among aircrews or tactical combat teams when planning and executing missions? These are a few challenging questions that lend themselves to investigation using behavioral criteria.

The need for a linkage with established theories in social psychology. Though LOC is hardly a new concept, its application to aviation has been recent. Research employing the new instruments such as the ASLOC and New-HAS should not be limited to concurrent validation on military aviators. Correlational research does not teach us much about the process. The research has compared LOC scores to scores on other measures of control and risk taking, notably old and new version of the HAS, and various self-report questionnaires (e.g., HES). This research may not progress much further than this, for lack of a theoretical framework. There is much more that we do not know concerning the dynamics and maintenance of beliefs concerning personal control and risk taking. New areas should be explored, which can link hazardous attitudes and LOC to behavioral and attributional variables that would serve as means of refining the construct. A rich heritage of research in applied social psychology exists (e.g., attributional biases, maintenance of the self-concept, cognitive consistency, risk taking) that is eminently applicable to aviation. The conceptual similarity between internal-external loci of control and cause, previously discussed, provides further evidence of the close correspondence between LOC and attribution theory and suggests that an integration of the two would be beneficial to the discipline of applied social psychology.

Stability and change over time; a need to understand the process. Previous research does not tell us much about how the need for control is maintained over time, especially in the face of repeated success and failure. This would necessitate a theoretical underpinning and a return to Rotter's social learning-expectancy theory. Unfortunately, many researchers seem to have drifted away from the close relationship between LOC and expectancy theory. In order to demonstrate hypotheses derived from this theory, criterion measures other than self-reports (i.e., behavioral measures) are needed. Instead of factor-analytic and attitudinal studies, it would be worthwhile examining in situ behavioral differences between internals and externals. A positive outcome of this research would be additional insight into the relationship between LOC and recent personal control-based expectancy theories, which may be better suited for examining the processes mediating changes in expectancies concerning one's control over the environment. It could answer questions about how behavioral outcomes and their valences affect the maintenance of a sense of personal control over time, and whether a sense of control can become a sense of overconfidence.

Perceived Personal Control and Risk Taking

Effects of experience on risk taking. It is clear that attitudes toward control over one's environment are related to risk taking preferences and overt behaviors. Their effects on the risk taking behaviors of military aviators at different stages of their careers are not clear. Are old pilots not bold pilots due to object lessons from personal bad experiences or from observing the misfortunes of

others? Alternatively, can we say that young, inexperienced pilots are more prone to risk taking than their older and supposedly more mature counterparts? Are differences in risk taking and risk management by aviators related to any of the hazardous attitudes, and if so, to which ones (e.g. antiauthority, macho, or invulnerable)? It would be worthwhile knowing the trends in personal control among aviators with different levels of experience. This could be ascertained by administration of Hunter's ASLOC scale, but it would only provide a partial answer to the question. We would not learn the degree to which individual differences in ASLOC scores are influenced by situational factors.

Cross-sectional and longitudinal trends in sense of personal control and risk orientation. Instruments that purport to measure LOC and hazardous attitudes have recently been psychometrically refined. Replication on larger samples of aviators would determine the extent to which the relationships found are stable and reproducible. Likewise, a representative sample of military aviators would demonstrate any consistent and stable differences that distinguish this group from the general and commercial aviation samples. One question raised by this review is whether combat experience changes one's perception of personal control when facing hazardous situations. The relationship between flight experience, age, and sense of personal control could also be investigated. This may answer some questions about LOC changes in response to aging (at what point in the life cycle does one become more [or less] cautious?).

Cross-sectional research has one major limitation: the confounding of life cycle and generational changes over time. Although it would be highly desirable to perform a longitudinal study over the career cycle of pilots, this would be very difficult in the absence of comprehensive tracking systems. Instead, researchers will have to make do with cross-sectional comparisons among cohorts. Sense of personal control can be measured as an individual difference variable by administering the ASLOC. Similarly, sense of personal control and optimistic bias can be measured as situationally determined attributional styles via questionnaires developed by Weinstein (1980). Similar to other studies of accident involvement, self-reports such as Hunter's HES, as well as archival data, can be used. Chapin (2001) points out that the optimistic bias, though quite robust, may not be universal. To some degree, optimism grows with successful experience, but this can be tempered by the knowledge of negative outcomes.

The study of LOC and risk orientation serves one simple goal: to make flying safer. This paper has delved into the research literature, attempting to find commonalities and differences in the findings, and has endeavored to understand the cognitive processes mediating risk perception and risk taking by aviators. Gaps in the knowledge base have been identified, and new directions in research suggested. Although LOC, HAS, and attribution theory are not new, there is a lot of research that remains to be done, before we really understand the dynamics of risk perception and risk management among aviators.

References

Abelson, R. P. (1959). Modes of resolution of belief dilemmas. *Journal of Conflict Resolution, 3*(4)*,* 343-352.

Alloy, L. B., Abramson, L. Y., & Viscusi, D. (1981). Induced mood and the illusion of control. *Journal of Personality and Social Psychology, 41*(6), 1129-1140.

Anderson, A., Hattie, J., & Hamilton, R. J. (2005). Locus of control, self-efficacy, and motivation in different schools: Is moderation the key to success? *Educational Psychology, 25*(5), 517-535.

Arthur, W. A., Barrett, G. V. & Alexander, R. A. (1991). Prediction of motor vehicle involvement: A meta-analysis. *Human Performance, 4*(2), 89-105.

Berlin, J. I., Gruber, E. V., Holmes, C. W., Jensen, P. K., Lau, J. R., Mills, J. W., & O'Kane, J. M. (1982). *Pilot judgment training and evaluation* (DOT/FAA/CT-82/56). Daytona Beach, FL: Embry-Riddle Aeronautical Universit

Cattell, R. B., Eber, W. H., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing.

Chapin, J. (2001). Self-protective pessimism: Optimistic bias in reverse. *North American Journal of Psychology, 3*(2), 253-262.

Collins, B. E. (1974). Four components of the Rotter internal-external scale. *Journal of Personality and Social Psychology, 29*(3), 381-391.

Duke, M., Johnson, T. C., & Nowicki, S. (1977). Effects of sports fitness camp experience on locus of control orientation in children ages 6 to 14. *The Research Quarterly, 48*(2)*,* 280-283.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69-106.

Federal Aviation Administration (1991). *Aeronautical decision making*. U.S. Department of Transportation, Federal Aviation Administration Advisory Circular 60-22.

Heider, F. (1958). The psychology of interpersonal relations. New York: John Wiley and Sons.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review, 5*(1), 74-95.

Holt, R. W., Boehm-Davis, D. A., Fitzgerald, K. A., Matyuf, M. M., Baughman, W. A., & Littman, D. C. (1991). Behavioral validation of a hazardous thought pattern instrument. *In Proceedings of the Human Factors Society 35th Annual Meeting,* (pp. 77-81). Santa Monica, CA: Human Factors Society.

Hunter, D. R. (1995, October). *Airman research questionnaire: methodology and overall results* (*Rep. No.* DOT/FAA/AM-95/27). Washington, D.C.: Federal Aviation Administration.

Hunter, D. R. (2002). Development of an aviation safety locus of control scale. *Aviation, Space, and Environmental Medicine, 73*(12), 1184-1188.

Hunter, D. R. (2005). Measurement of hazardous attitudes among pilots. *International Journal of Aviation Psychology, 15*(1), 23-43.

Hunter, D. R. (2006). Risk perception among general aviation pilots. *International Journal of Aviation Psychology, 16*(2)*,* 135-144.

Hunter, D.R. (2008). [Military ASLOC data analysis]. Unpublished data analysis.

Jensen, R. S., & Benel, R. A. (1977). *Judgment evaluation and instruction in civil pilot training.* (FAA-RD-78-24). Savoy, IL: University of Illinois Aviation Research Laboratory.

Jones, E. E., & deCharms, R. (1958). The organizing function of interaction roles in person perception. *Journal of Abnormal and Social Psychology, 57*(2)*,* 155-164.

Jones, J. W., & Wuebker, L. (1985). Development and validation of the safety locus of control scale. *Perceptual and Motor Skills, 61*(1), 151-161.

Joseph, C., & Ganesh, A. (2006). Aviation safety locus of control in Indian aviators. *Indian Journal of Aerospace Medicine, 50*(1), 14-21.

Klein, C. T. F., & Helweg-Larsen, M. (2002). Perceived control and the optimistic bias:A meta-analytic review. *Psychology and Health, 17*(4), 437-446.

Lerner, M. J. (1997). What does the belief in a just world protect us from: The dread of death or the fear of undeserved suffering? *Psychological Inquiry, 8*(1), 29-32.

Lerner, M. J. (2003). The justice motive: Where social psychologists found it, how they lost it, and why they may not find it again. *Personality and Social Psychology Review, 7*(4), 388-399.

Lerner, M. J., & Miller, D. T. (1978). Just world research and attribution process: Looking back and ahead. *Psychological Bulletin, 85*(5)*,* 1030-1051.

Lester, L. F., & Bombaci, D. H. (1984). The relationship between personality and irrational judgment in civil pilots. *Human Factors*, *26*(5), 565-572.

Lester, L. F., & Connolly, T. J. (1987). The measurement of hazardous thought patterns and their relationship to pilot personality. *Proceedings of the Fourth International   Symposium on Aviation Psychology*, 286-292.

Lynch, M. L., Ogg, W. D., & Christensen, M. G. (1975). Impact of a life planning workshop on perceived locus of control. *Psychological Reports, 37*(3 PT 2), 338-343.

Montag, I., & Comrey, A. L. (1987). Internality and externality as correlates of involvement in fatal driving accidents. *Journal of Applied Psychology, 72*(3), 339-343.

Perloff, L. S., & Fetzer, B. K. (1986). Self-other judgments and perceived invulnerability to victimization*. Journal of Personality and Social Psychology, 50*(3), 502-510.

Phares, E. J. (1976). *Locus of control in personality*. Morristown, N. J.: General Learning Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.) *Advances in experimental social psychology* (vol.10, pp.173-200). New York: Academic Press

Rotter, J. B. (1954). *Social learning and clinical psychology*. Englewood Cliffs, N.J.: Prentice-Hall.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs 80* (1 Whole No. 609).

Stewart, J. E. (2006). *Locus of control, attribution theory, and the "five deadly sins" of aviation. (Tech. Rep. No. 1182)*. Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Twenge, J. M., Zhang, L., & Im, C. (2004). It's beyond my control: A cross-temporal meta-analysis of increasing externality in locus of control, 1960-2002. *Personality and Social Psychology Review, 8*(3), 308-319.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5)*,* 806-820.

Weinstein, N. D. (1987). Unrealistic optimism about the susceptibility to health problems: Conclusions of a community-wide sample. *Journal of Behavioral Medicine, 10*(5), 481-500.

Weinstein, N. D., & Klein, W. M. (1996). Unrealistic optimism: present and future. *Journal of Clinical and Social Psychology, 15*(1), 1-8.

Wetmore, M., Bos, P., & Lu, C. (2007). Conflict management strategies in the flight training environment. *International Journal of Applied Aviation Studies, 7*(2). 256-271.

Wichman, H., & Ball, J. (1983). Locus of control, self-serving biases, and attitudes toward safety in general aviation pilots. *Aviation, Space, and Environmental Medicine, 54*(6), 507-510.

Wilson, D. R., & Fallshore, M. (2001). Optimistic and ability biases in pilots' decisions and perceptions of risk regarding VFR flight into IMC. *Proceedings of the 11th International Symposium on Aviation Psychology*, Columbus, OH.

Collegiate Flight Training:
Making Progress in the Face of Adverse Conditions

Beth M. Beaudin-Seiler
Western Michigan University
College of Aviation
237 North Helmer Road
Battle Creek, MI, 49037
269-964-4579
beth.seiler@wmich.edu

Jeffrey M. Beaubien
Cognitive Systems Engineering Division,
Aptima, Inc.
Woburn, MA

Ryan C. Seiler
Western Michigan University,
College of Aviation
Kalamazoo, MI

Abstract

This paper describes the approach that one flight program took to better track the progress of its student pilots. The project's goals were to identify the effects of training gaps, evaluate the number of semesters to complete a private pilot certificate, evaluate the quality of instruction provided to students, determine when students should be flagged for lack of progress, and suggest remediation strategies. Multiple regression analysis was used to assess the effects of training gaps and instructor quality on the number of semesters required to complete the private pilot's certificate. Results show that gaps in training explain significant criterion variance even when controlling for other relevant variables. Newly developed tools, such as the Gaps in Instruction Adjustment Matrix, may help to standardize the administrative decisions concerning the amount of remedial training required following a gap in instruction.

Collegiate Flight Training:

Making Progress in the Face of Adverse Conditions

As the aviation industry continues to hire, collegiate flight training programs are struggling to meet the ever-increasing demand for qualified airline pilots. According to Federal Aviation Administration (FAA) forecasts, the number of passengers carried by U.S. commercial air carriers will increase by 2.8% in 2008 alone and is on track to hit one billion by the year 2015 (FAA, 2007). At the same time, the Bureau of Labor Statistics (BLS) indicates there will be a 13% increase in demand for airline pilots through 2016. This translates into approximately 27,600 more airline pilots than currently exist today (BLS, 2008).

Because their fates are so highly intertwined, factors that prevent collegiate aviation programs from successfully graduating a sufficient number of qualified pilots will likely have a ripple effect throughout the larger air transportation industry. With this in mind, this exploratory study was conducted by Western Michigan University's College of Aviation (WMU) to understand some of the challenges that students encounter during their private pilot training course. Emphasis focused on the number and types of gaps in instruction, as well as the quality of instruction that is provided to students.

Factors that Influence Success in Flight Training
Just some of the factors that can influence success in flight training have been highlighted.

Emotional Maturity. Being a new college student brings with it a variety of challenges. These include being away from home for the first time, anxiety about making decisions on their own, and handling the pressures of their chosen career path (Tokar, Withrow, Hall, & Moradi, 2003). The emotions of these young, immature students affect not only what they think about, but also how they process information (Tiedens & Linton, 2001). As such, these factors can greatly influence the students' education-related activities such as their study habits, their decisions to attend (or not attend) class, and their decisions to show (or not show) for flight slots.

Intelligence. Collegiate aviation programs often admit students who may not have prior flight experience or exceedingly high levels of general cognitive ability. According to a study of Air Force trainees (Ree, Carretta, & Teachout, 1995), selecting applicants with high levels of intelligence and prior job knowledge should lead to better performance in training, and by extension, on the job. However, if there is a scarce population of potential applicants, intelligence tests should take precedence over tests of prior job knowledge (Ree et al., 1995). Previous research confirms that measures of intelligence, which include elements of general cognitive ability, verbal ability, quantitative ability, and mechanical ability, are effective predictors of job performance across a wide range of jobs including those in commercial and military aviation (Hunter & Burke, 1994; Ree et al., 1995).

Decay of Skills. In dealing with collegiate flight students, it is not always possible to keep them focused on their flying skills. Other classes, weather delays, and personal issues can cause delays in their training. Research done by Healy,

Wohldmann, Parker, & Bourne (2005) found that participants who experienced a one-week delay in conducting new tasks seemed to have a lack of transfer from the original task learned. However, they also noted that by the end of retraining, participants were able to perform the secondary tasks equivalent to those participants who had only performed the original task, which indicates participants can overcome the transfer issue with sufficient additional practice. (Healy, et al. 2005) For a flight student that means additional flights in order to get back up to speed.

Credentials of the Instructor. Flight instructors are not like classroom instructors, flight instructors only need to pass the Federal Aviation Association's Certified Flight Instructor license and they are approved to teach others how to fly. Compared with elementary and high school instructors who are given incentives to become Nationally Board Certified (Stronge, et al, 2007) or collegiate instructors that hold PhD level degrees, it is easy to see the difference in levels between "instructors". Research has shown that instructor quality is most often associated with gains in student performance. (Hertert & Teague, 2003).

Statement of the Problem

Western Michigan University's flight science program mostly accepts traditional college age students, those whom have recently graduated from high school. The program also does not currently screen candidates based on their level of general cognitive ability. Therefore, it is imperative that the program be able to assess the capabilities of its students early during their academic careers, identify struggling students, and develop remediation plans that are fair and equitable to both the student and the program.

At the personal request of the Dean of the College of Aviation, the research team began to explore factors that affect progress (or lack thereof) in WMU's flight science program, but until now were not previously explored in a systematic fashion. These factors included: gaps in training that are caused by unfavorable weather patterns, semester breaks, holiday closures, and the like; high levels of turnover among the cadre of flight instructors; the decreased experience level among the new flight instructors; and the number of semesters it takes for the students to complete the private pilot course. With this in mind, the research team culled through student archival records to identify those factors that explain both statistically and practically significant variance in the amount of time required to complete the private pilot certificate.

Method

Participants

All participants were enrolled in the Private Pilot Course at WMU's College of Aviation during the Fall 2003 and Fall 2005 college semesters. (Note: 2004 data was not included as the College converted its fleet of aircraft during that year.) None had previously obtained a private pilot certificate prior to beginning their course work. While some may have had flight experience prior to starting the private pilot curriculum at WMU, it was not significant enough to warrant any

credit given towards their private pilot certificates. There were 49 participants: 23 students during the Fall 2003 semester and 26 students during the Fall 2005 semester. Consistent with previous years' enrollment records, most of the participants were male (95.8% in the Fall 2003 semester, and 88.4% in the Fall 2005 semester). The mean age of the participants was 20.35 years (SD = 1.49), which is again consistent with previous enrollment records.

Flight Training Curricula

The Fall 2003 and Fall 2005 semesters were similar in many respects, but they were not identical. The 2003 private pilot curriculum had the following objectives and standards:

Objectives. During this stage, the student shall complete all aeronautical experience, skill and knowledge requirements to accomplish all private pilot areas of operation and flight tasks. In addition, the student shall be introduced to elements of crew resource management in flight operations and/or flight training device applications, and demonstrate proper flight ethics and responsibility.

Completion Standard. The student shall complete all private pilot tasks to practical test standards. The student shall complete all lessons and final stage checks to specified performance standards. The student shall obtain the private pilot certificate.

Aeronautical Experience. This course includes 40 hours of flight training for private pilot certification. These 40 hours include a maximum of 2 hours of flight training device (FTD) training time. The curriculum solo would take place at Lesson 11, with at least 10.6 hours of flight time being completed (Western Michigan University, 2002).

By way of comparison, the 2005 private pilot curriculum had the following objectives and standards:

Prerequisite Experience. A student must hold a recreational or student pilot certificate prior to enrolling. A Second-Class aviation medical certificate is required. Previous flight experience is not required.

Objectives. During the course, the student shall complete all aeronautical experience, skill and knowledge requirements to accomplish all private pilot areas of operation and flight tasks. In addition, the student shall be introduced to elements of Crew Resource Management in flight operations and/or flight training device (FTD) applications. In addition, they will demonstrate proper flight ethics and responsibility.

Completion Standards. The student shall complete all private pilot tasks to practical test standards. The student shall complete all lessons and the final stage check to the specified performance standards. The student shall obtain the private pilot certificate.

Aeronautical Experience. This course includes a minimum of 50 hours of flight training for private pilot certification. These 50 hours (with a possibility of 10 hours

of flex time if needed) include a maximum of 3 hours of FTD training time. The curriculum solo would take place at Lesson 13, with at least 15.3 hours being completed (Western Michigan University, 2004).

The objectives and completion standards for these two curriculums were exactly the same, the only difference being the number of hours in the aeronautical experience category. In 2003, the curriculum called for 40 hours with 2 hours in the simulator and solo at Lesson 11. In 2005, the curriculum called for 50 hours with 3 hours in the simulator and solo at Lesson 13. The 2005 curriculum also allowed for 10 hours of "flex time" which allows the student an additional 10 hours anywhere in the curriculum where extra sorties may be needed. This means a student could complete the private course in 40 hours and use those 10 hours of "flex time" in another part of the curriculum. Alternatively, the student may need the extra sorties during the private curriculum and according to the syllabus, could use that 10 hours bringing them to 50 hours for the private pilot course.

Design and Procedure
Archival Records. Most of the data were collected from archival sources, including student log books and flight records. Data collection occurred after the participants had completed the private pilot theory and flight courses. As such, there was no real-time communication with either the participants or their flight instructors with regard to their progress in the private pilot course.

Measured variables included: hours to solo; hours to private rating; number of instructor changes; number of days consecutively not flown during the first half of the curriculum (meaning the first 50% of the total lessons in the private pilot syllabus); number of consecutive days not flown during the second half of the curriculum (meaning the second 50% of the total lessons in the private pilot syllabus); total number of gaps during training; length of each gap during training; and total number of semesters required to complete the private pilot certificate. Additional variables (described below) were based on the raw scores and flight instructor input that was collected during a series of focused group interviews.

Adjusted Total Time. Regrettably, many students experience gaps in their professional training. Some of these gaps are pre-planned. These include gaps caused by semester breaks, holiday closures, and summer vacations. Other gaps, however, are unanticipated. These include gaps caused by inclement weather, military commitments, and turnover among flight instructors. Regardless of the cause, the flight curriculum was not written to accommodate gaps in training. Historically, it has been left to the Chief Flight Instructor to determine how much additional flight training is required to offset these gaps. To standardize the determination process, a *Gaps in Training Adjustment Matrix* was developed (see Appendix A). This matrix was derived from hours of focus groups conducted with subject matter experts, specifically the Chief, the Assistant Chiefs, and the Lead Faculty Flight Instructors to arrive at the number of hours reasonably needed to regain proficiency. Each student's adjusted total time was calculated as their total number of flight hours minus the number of remedial hours that were recom-

mended from the matrix. The adjusted total time is believed to be a more accurate estimate of the student's actual proficiency, because it takes into account the effects of skill decay that are caused by gaps in instruction. All source data used to calculate the students' adjusted total time were culled from their log books and flight records as described previously.

Quality of Instruction. As the airlines hire, it has become difficult for collegiate aviation programs to retain qualified flight instructors because they can command much higher salaries flying commercially. As a result, the current average tenure of certified flight instructors at WMU is at a historical low. At the same time, there is also a small cadre of highly-trained flight instructors who choose not to fly commercially but instead focus their efforts on teaching. Because it is not possible to compare the quality of instruction provided by these two cadres of flight instructors directly, a *Quality of Instruction* formula was developed to address the quality of instruction being given to students (see Appendix B). This formula was derived from many hours of focus groups with subject matter experts, namely the Chief, the Associate Chiefs, and the Lead Faculty Flight Instructors to determine the appropriate weighting and distribution of the formula.

Results

Descriptive Statistics. A review of the descriptive statistics suggest that students enrolled in the Fall 2005 semester took longer to complete the private pilot course (M = 70.0, SD = 16.36) than students in the Fall 2003 semester (M = 64.78, SD = 9.52). However, the difference was not statistically significant, $t(47) = 1.36$, p = .19, thereby suggesting that the two groups are comparable. The descriptive statistics also suggest that the quality of instruction differed between the two semesters. Specifically, instructors from the Fall 2003 semester were rated as more proficient (M = 2.59, SD = 1.60) than their Fall 2005 counterparts (M = 1.31, SD = .55). These results were statistically significant $t(47) = 3.84$, p = .00. This latter finding is likely due to the fact that the aviation industry had started hiring again and the experience of flight instructors was not as deep as during 2003 when the industry was still coping with downsizings and furloughs.

Data Quality Analysis. Hours to solo and hours to private pilot certificate were found to be moderately correlated ($r = .537$, p = .00). This degree of inter-correlation is moderate and should not systematically bias the regression weights during subsequent statistical analyses. However, hours to private pilot certificate and adjusted total hours were highly correlated ($r = .729$, p = .00). This is only natural, as the adjusted total hours are based on the raw hours to private pilot certificate. This high degree of inter-correlation will likely bias the regression weights during subsequent analyses if both variables are used during the same analysis.

Inferential Statistics. A series of hierarchical multiple regression analyses (Cohen & Cohen, 1983) were conducted to assess the effects of gaps in training on the number of semesters required to complete a private pilot certificate. In all cases, semester (Fall 2003 or Fall 2005) and instructor quality ratings were entered in the first block to control for differences in the quality of instruction provided to students. Next, the students' number of flight hours (hours to solo and hours to private pilot certificate) was entered as a block to control for differences in student

experience. Finally, the effect of gaps in instruction was entered to assess the effects of skill decay on learning outcomes. Because there was no a priori reason to believe that any one method of operationalizing gaps in instruction was better (from the perspective of statistical prediction) than any other, the researchers ran four separate regression analyses to compare their relative effects. Specifically, the researchers operationalized the effect of gaps in instruction as the maximum number of gaps, the average number of gaps, the total number of gaps, and the adjusted total time. The results are summarized in Tables 1 through 4 below. Readers should note that the sample sizes vary because of missing data.

Table 1
Operationalization #1: Maximum Number of Gaps in Instruction

Step	Variable	b	Std. Error	β	Sig.	R^2 Change	Sig. F Change
1	Instructor Quality	-.056	.090	-.098	.536	.132	.039
	Semester	.457	.236	.305	.059		
2	Hours to Solo	-.019	.035	-.095	.585	.239	.001
	Hours to Private	.031	.009	.554	.002		
3	Maximum Gaps (First Half)	.018	.006	.328	.003	.313	.000
	Maximum Gaps (Second Half)	.012	.002	.474	.000		

Note 1: n = 48
Note 2: Dependent Variable = Number of Semesters to Complete Private Pilot
Note 3: Actual R^2 = .683 (Adjusted R^2 = .638)

Table 2
Operationalization #2: Average Number of Gaps in Instruction

Step	Variable	b	Std. Error	β	Sig.	R^2 Change	Sig. F Change
1	Instructor Quality	-.095	.187	-.093	.615	.042	.495
	Semester	.256	.313	.150	.420		
2	Hours to Solo	-.022	.042	-.106	.611	.312	.002
	Hours to Private	.034	.011	.649	.003		
3	Average Gaps (First Half)	.022	.015	.226	.151	.136	.032
	Average Gaps (Second Half)	.031	.013	.336	.019		

Note 1: n = 35
Note 2: Dependent Variable = Number of Semesters to Complete Private Pilot
Note 3: Actual R^2 = .490 (Adjusted R^2 = .384)

Table 3
Operationalization #3: Total Number of Gaps in Instruction

Step	Variable	b	Std. Error	β	Sig.	R² Change	Sig. F Change
1	Instructor Quality	-.056	.090	-.098	.536	.132	.039
	Semester	.457	.236	.305	.059		
2	Hours to Solo	-.019	.035	-.095	.585	.239	.001
	Hours to Private	.031	.009	.554	.002		
3	Total Gaps	.149	.022	.654	.000	.324	.000

Note 1: n = 48
Note 2: Dependent Variable = Number of Semesters to Complete Private Pilot
Note 3: Actual R^2 = .695 (Adjusted R^2 = .660)

Table 4
Operationalization #4: Adjusted Total Time

Step	Variable	b	Std. Error	β	Sig.	R² Change	Sig. F Change
1	Instructor Quality	-.056	.090	-.098	.536	.132	.039
	Semester	.457	.236	.305	.059		
2	Hours to Solo	-.019	.035	-.095	.585	.239	.001
	Hours to Private	.031	.009	.554	.002		
3	Adjusted Total Time	-.061	.008	-.968	.000	.378	.000

Note 1: n = 48
Note 2: Dependent Variable = Number of Semesters to Complete Private Pilot
Note 3: Actual R^2 = .748 (Adjusted R^2 = .719)

From these analyses, we can deduce several important lessons. First, gaps in instruction, regardless of how they are operationalized, clearly affect the number of semesters that are required to obtain a private pilot's license. In all cases, the R^2 change values were both large (ranging between 13.6% and 37.8% incremental criterion variance) and statistically significant. The smallest effect was observed for operationalization #2, the average number of gaps in instruction, and the largest was observed for operationalization #4, the matrix-adjusted total time. In all cases, however, these effects held even after controlling for both the quality of instruction given (instructor quality, semester) and the student's level of experience (hours to solo, hours to private).

Second, the various ways of operationalizing the gaps in instruction produce somewhat different results. Specifically, the average number of gaps in instruction only explained 38.4% of the variance in the number of semesters required to complete the private pilot certificate. By way of comparison, the maximum and total number of gaps in instruction explained both significantly more variance (68.3% and 66.0%, respectively) in the number of semesters that are required to obtain a private pilot's license.

Finally, the adjusted total time (which was calculated using the *Gaps in Instruction Adjustment Matrix*) explained the most criterion variance (71.9% of the variance in the number of semesters required to obtain a private pilot's license). This held true even when the uncorrected, actual number of hours to private were

already included in the analysis (see Table 4). It should be noted that these two variables were highly correlated ($r = .729$), which accounts for the negative regression weight ($\beta = -.968$) for the adjusted total time variable. Despite the reversed sign, the explanatory power for this operationalization suggests that that the statistical correction generated by using the *Gaps in Instruction Adjustment Matrix* is working as intended. Based on the pattern of results in Tables 1 through 4, we believe that statistically adjusting for the gaps in training using the matrix may allow for a better assessment of student progress than simply considering the number or length of gaps themselves. These results are obviously very preliminary and need to be cross-validated with independent samples of student pilots.

Discussion

The purpose of this study was to assess the effects of instructor quality and gaps in training on one critical learning outcome: the number of semesters required to complete private pilot certification. In recent years, this outcome has become increasingly important as the airlines hire to meet projected demand for air travel. As noted earlier, factors that prevent collegiate aviation programs from successfully graduating a sufficient number of qualified pilots will have a ripple effect throughout the larger air transportation industry.

Our analyses included several factors for which the university has absolutely no control over: the quality of instructors, and gaps in instruction that are caused by weather, university breaks, and students' military commitments. The analysis also included some elements that the university can control, albeit imperfectly: statistically adjusting the students' hours to compare their progress to that of a curriculum, which also does not account for gaps in training. Doing so allows the university to assess the students' progress more accurately. The next step is to use these analyses to help determine when students should be "flagged" for lack of progress, and to suggest remediation strategies so that successful learning outcomes can be achieved.

Now that we have provided some initial empirical evidence to support the use of the *Gaps in Instruction Adjustment Matrix*, it will serve as a foundation for developing fair and equitable remediation plans that can be put in to practice. For example, no longer will students face arbitrary decisions from the Chief Flight Instructor as to how many remediation flights are necessary after returning from spring break. Also, by understanding that the longer it takes to solo a student is positively correlated with the amount of time required to earn his/her private pilot certificate, we can now build benchmarks into the pre-solo stage that allows for monitoring of students' performance later on. Using this data, we will be able to catch the ones who are struggling earlier, adjust their times to better compare their flight hours against the syllabus, and make intelligent instructor changes – for example, assigning more seasoned instructors to struggling students rather than simply assigning instructors based on availability – that will hopefully increase their chances of obtaining their private pilot certificate on schedule. Below are two examples of how we hope to apply the findings from this study in our administrative decisions going forward.

Two Examples

   Student A. Student A is a new private pilot student with zero outside flight experience. She has a newly certified flight instructor and she has hit all the pre-solo and solo milestones on course with the syllabus outline. However, as she nears the first progress checkpoint, she begins to have some trouble. She begins struggling with the lessons, feels inadequately prepared for the learning objectives covered, and needs to repeat lessons. After three attempts at a single lesson, she is flagged by the system and is brought to the attention of a Lead Flight Instructor. This Lead Flight Instructor observes the progress Student A has made throughout the private pilot curriculum and sees that she has 35 hours at the point of the first progress check. Comparing her to the syllabus, he sees that at the point of the first progress check she should be at approximately 23 hours. He then adjusts her time for any gaps in training using the *Gaps in Instruction Adjustment Matrix*, and notes that in reality she is at about 26 hours. He then calculates the quality of instruction that she is currently being given using the *Quality of Instruction Formula* and sees that her instructor rating is low (2.0 out of 5.0). Based on this information, he decides not to implement an instructor change, but rather sits down with both student and instructor to identify the best plan for moving forward.

   Student B. Student B is a new private pilot student with zero outside flight experience. He has a newly certified flight instructor. He has not hit any pre-solo milestones and this has been brought to the attention of a lead flight instructor on several occasions. The Lead Flight Instructor observes the progress Student B has made throughout the private pilot curriculum, and sees that he has 30 hours and has not completed his solo. Comparing him to the syllabus, he sees that at the point of solo the student should be at approximately 15 hours. He finds no gaps in training and therefore no adjustments need to be made to the student's time. The lead flight instructor then calculates the *Quality of Instruction* being given and can see that his instructor rating is again low (1.5 out of 5.0). Based on this information, he decides to implement an instructor change and purposefully selects an instructor with a rating of a 3.0 or higher.

Directions for Future Research

   Our next step will be to devise an appropriate remediation policy that outlines expectations of student's progress, along with remediation plans if progress is not made. In addition, it is important to continue our research on the *Gaps in Training Adjustment Matrix* and *Quality of Instruction Formula* to ensure that we can provide fair and equitable remediation solutions for students in the private pilot curriculum. Additional research also needs to be completed in the instrument, commercial, and multi-engine courses, where similar remediation plans can be outlined and implemented.

References

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.Federal Aviation Administration. (2007). *FAA forecasts steady growth in air travel demand*. Washington, DC. Retrieved February 25, 2008, from http://www.faa.gov/news/press_releases/news_story.cfm?newsId=8358.

Healy, A.F, Wohldmann, E.L., Parker, J.T., Bourne, Jr., L.E (2005). Skill training, retention, and transfer: The effects of a concurrent secondary task. *Memory & Cognition*, 33(8), 1457-1471. Retrieved September 18, 2008, from Research Library database. (Document ID: 1020189231).

Hertert, L., Teague, J. (2003). Narrowing the Achievement Gap: A Review of Research, Policies and Issues. Report. EdSource. (ERIC Document Reproduction Service No. ED473724)

Hunter, D. & Burke, E. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *International Journal of Aviation Psychology, 4*, 297-313.

Ree, M., Carretta, T., & Teachout, M. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80*, 721-730.

Stronge, J.H, Ward, T.J, Tucker, P.D., Hindman, J.L., McColsky, W., Howard, B. (2007). *National Board Certified Teachers and Non-National Board Certified Teachers: Is There a Difference in Teacher Effectiveness and Student Achievement? Journal of Personnel Evaluation in Education, 20(3-4), 185-210. Retrieved September 17, 2008, from ABI/INFORM Global database. (Document ID: 1455674221).*

Tiedens, L. & Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social   Psychology, 81*, 973-988.

Tokar, D., Withrow, J., Hall, R., & Moradi, B. (2003). Psychological separation, attachment security, vocational self-concept, crystallization, and career indecision: A structural equation analysis. *Journal of Counseling Psychology, 50*, 3-19.

U.S. Department of Labor, Bureau of Labor Statistics. (2008). *Aircraft Pilots and Flight Engineers*. Washington, DC. Retrieved March 18, 2008, from http://www.bls.gov/oco/pdf/ocos107.pdf.

Western Michigan University. (2002). *2003 Professional Flight Curriculum.* Unpublished curriculum. Approved by Grand Rapids, MI Flight Standards District Office (FSDO). Battle Creek, MI: Author.

Western Michigan University. (2004). *2005 Professional Flight Curriculum.* Unpublished curriculum. Approved by Grand Rapids, MI Flight Standards District Office (FSDO). Battle Creek, MI: Author.

The *Gaps in Training Adjustment Matrix* was developed by interviewing the Chief Flight Instructor, the Assistant Chief Flight Instructor, and several Lead Flight Instructors, many of whom hold Master or Gold Seal certified flight instructor (CFI) certifications. The matrix was designed to predict the likely number of remedial flight hours required to bring the student back up to the same level of proficiency that they had before the gap occurred. Gaps were defined as a period where the student was not in the aircraft or the flight-training device (FTD) for more than 7 consecutive days. To use the matrix effectively, one must know only two variables: where in the curriculum the gap occurred, and the length of the gap in instruction. The matrix is shown in Table A1 below.

Table A1
Gaps in Training Adjustment Matrix

First Half of Curriculum		
Length of Gap in Instruction	Additional Flights Required	Not to Exceed _X_ Hours
7 to 16	1	2
17 to 26	2	4
27 to 36	3	6
37 to 46	4	8
47 to 56	5	10
57 to 66	6	12
67 to 76	7	14
77 to 86	8	16
87 to 96	9	18
Over 97	10	20
Second Half of Curriculum		
Length of Gap in Instruction	Additional Flights Required	Not to Exceed _X_ Hours
7 to 16	1	1
17 to 26	2	2
27 to 36	3	3
37 to 46	4	4
47 to 56	5	5
57 to 66	6	6
67 to 76	7	7
77 to 86	8	8
87 to 96	9	9
Over 97	10	10

Consider, for example, a student enrolled in his/her first half of the private pilot curriculum. If that student had a gap in instruction which lasted 9 consecutive days, that student would need one additional flight not to exceed 2 hours in duration. If that same gap occurred during the second half of the curriculum, that same student would need only one additional flight totaling 1 hour in duration. From the data depicted in Table 5, it is clear that the location of the gap (during the first or second half of the curriculum) is critical to determining the amount of remediation required. Specifically, if the gap occurred during the first half of the curriculum, more remediation would be required than if the gap occurred during the second half.

Appendix B

Development of the *Quality of Instruction Formula* followed the same manner as the *Gaps in Training Adjustment Matrix*. Again, the lead author conducted a series of focus group interviews with the Chief Flight Instructor, the Assistant Chief Flight Instructor, and several Lead Flight Instructors, many of whom hold Master or Gold Seal certified flight instructor (CFI) certifications. The *Quality of Instruction Formula* was designed to statistically control for differences in the quality of instruction that is provided to the students. To calculate the quality of instruction effectively, one must know two variables: The instructor's pass rate on private pilot certificates and the instructor's total number of hours of dual given instruction. For the purpose of this study, the *Quality of Instruction Formula* was operationalized as:

Quality of Instruction = (Pass Rate x .30) + (Dual Given Instruction Experience x .70)

Pass Rate on Private pilot certificates

90 - 100% = 5

80 - 89% = 4

70 - 79% = 3

60 - 69% = 2

Less than 59% = 1

Dual Given Instruction Experience

701 - 900 hours = 5

501 - 700 hours = 4

301 - 500 hours = 3

101 - 300 hours = 2

Less than 100 hours = 1

In the event that more than one instructor was involved during the student's private pilot curriculum, each instructor's score was weighted by the amount of instruction provided (total amount of instruction = 100%) and then summed across instructors.

Once the instructors in this study were identified and their *Quality of Instruction* ratings were independently calculated by the research team, the ratings were validated by correlating the data with independent (blind) ratings from the Lead Flight Instructors and Assistant Chief Flight Instructors. In making their ratings, the Lead Flight Instructors and Assistant Chief Flight Instructor used several criteria, including their experiences working with the instructors on a daily basis, seeing how they interact with their students, and observing the progress of their students. The mean correlation between the two sets of scores was $r = .88$, suggesting that the *Quality of Instruction* ratings were working as intended.

Classification and Analysis of Errors Reported in Aircraft Maintenance Manuals

Bonnie Lida Rogers
National Institute for Aviation Research
Wichita State University
Wichita, KS

Christopher J. Hamblin
National Institute for Aviation Research
Wichita State University
Wichita, KS

Address Correspondence to
Alex Chaparro
Department of Psychology
National Institute for Aviation Research
Wichita State University, Wichita, KS
alex.chaparro@wichita.edu
Phone: 1-316-978-3038
FAX: 1-316-978-3086

Abstract

Previous research has identified maintenance information as one of the primary causal factors of maintenance error. Incorrect maintenance information has also been cited as a contributing factor in a number of recent aircraft mishaps. To date no one has studied the types of errors found in aircraft maintenance manuals published by manufacturers. The purpose of this research is to analyze Publication Change Requests (PCRs) to document the most frequently reported types of errors found in aircraft maintenance manual, to identify how errors vary across Air Transport Association (ATA) chapters, and identify the corrective actions required to address the cited problem. The most common request was for additional procedural information followed by requests to add or change the language to improve clarity. The results show that the majority of PCRs (42%) cited procedures found in Chapters 27 (Flight controls), 32 (Landing gear), and 71 (Powerplant).

Classification and Analysis of Errors Reported
in Aircraft Maintenance Manuals

In 2003, Air Midwest Beechcraft 1900D with 19 passengers and 2 crew crashed shortly after takeoff. The National Transportation Safety Board accident investigation (NTSB, 2004b) revealed that the airplane's elevator control system was incorrectly rigged during a maintenance check restricting the airplanes elevator travel. The report cited Air Midwest's maintenance procedures and documentation as one of four probable causes of the accident. Eight months later, another fatal mishap involving a Beechcraft 1900D occurred due to an incorrectly rigged elevator trim. The National Transportation Safety Board determined (NTSB, 2004a) that the probable cause(s) of the accident were as follows:

> The improper replacement of the forward elevator trim cable, and subsequent inadequate functional check of the maintenance performed, which resulted in a reversal of the elevator trim system and a loss of control inflight. Factors were the flight crew's failure to follow the checklist procedures, and the aircraft manufacturer's erroneous depiction of the elevator trim drum in the maintenance (p 2).

The two accidents illustrate the critical role of maintenance practices and of the supporting maintenance documentation to flying safety. These maintenance related accidents are not isolated events. Analyses of major commercial aircraft accidents that occurred between 1959 and 1983 reveal that maintenance and inspection deficiencies account for 12% to 15% of commercial mishaps and are the fourth leading cause of accidents (Sears, 1986). Maintenance deficiencies account for a similar proportion (e.g., 17%) of naval aviation mishaps resulting in the loss of an aircraft or fatality (Ricci, 2003).

Causal Factors of Maintenance Error
Maintenance errors have a variety of causes and these causes can be organized into three broad categories. There are errors that 1) represent a failure to properly execute a correct plan of action; 2) errors resulting from the execution of an inadequate plan and 3) intentionally choosing a course of action that is a violation of formal rules and established procedures or that deviates from unofficial norms or standard practice (Reason & Hobbs, 2003). In the first case, an error may result from a misidentification of a signal or the failure to detect a defect during inspection. An Aviation Maintenance Technician (AMT) may misread an instrument or fail to detect a crack due to poor lighting or an interruption while performing a visual inspection. In the second case, errors typically arise from the misapplication of a useful rule or the application of a bad rule. For instance, a maintainer may develop a rule regarding what the standard torque values or tolerances for an aircraft component may be but fail to identify exceptions to the rule thereby leading to an error. Alternatively, a maintainer may adopt a habit that becomes part of their routine when performing a maintenance procedure but which has unintended consequences. In the late 1970's at one airline it became standard practice to use a forklift to support the engine/pylon assembly on DC-10s when replacing the assembly (NTSB, 1979). The use of the forklift in some cases caused unintended structural damage that resulted in a subsequent engine separation on takeoff and the loss of one aircraft, its passengers, and crew. Unlike the first two cases, a dis-

tinguishing characteristic of violations is that they are often intentional, not with the aim of bringing about bad consequences but rather to circumvent maintenance procedures. These procedures might be incorrect, lacking sufficient detail, unduly complicated, and/or burdensome. In addition, organizational and situational factors including short staffing, lack of appropriate tools, and situational considerations (i.e., on time departure and arrival) may predispose AMTs to engage in such behavior.

Several recent studies have sought to identify and classify the causal factors that contribute to maintenance error (A. Chaparro, Groff, Chaparro, & Scarlett, 2002; Hobbs & Williamson, 2003; Marx & Graeber, 1994; Patankar & Kanki, 2001; Ricci, 2003). Lattanzio, Patankar, and Kanki (2008) analyzed reports from AMTs submitted via the Aviation Safety Reporting System (ASRS) using MEDA (Maintenance Errors Decision Aid), a tool used to investigate and identify contributing factors to maintenance incidents (Rankin, Allen, & Sargent, 1998). The MEDA analysis identified 458 ASRS reports describing a procedural error that were defined as "information not understandable," "information incorrect," "information not enough," "information not used," and "information unavailable." They performed a content analysis of the 458 ASRS reports to identify and characterize the top 10% of reports that were most representative of the larger set. This analysis indicated that maintenance information (i.e., procedures in the Aviation Maintenance Manual (AMM), task cards, job cards, service bulletins, etc) was a significant causal factor of maintenance error. Table 1 shows the most commonly reported document deficiencies cited in the reports.

Table 1
The most frequently cited maintenance deficiencies cited in maintenance related ASRS reports.

Document Deficiencies	% (N=46)
Missing information	48
Incorrect Information	19
Difficult to interpret	19
Conflicting information	19

The results of Lattanzio et al. (2008) are in accord with other published findings. McDonald, Corrigan, Daly, & Cromie (2000) conducted a survey on the use of maintenance manual procedures as part of a larger study on organization aspects of safety. Thirty-four percent of their respondents reported performing routine maintenance tasks in ways different from those outlined in the documentation. The two most frequently cited reasons for not following the manual were that there was an *easier* and there was a *faster* method of performing the procedure. Similarly, a survey of Australian maintenance technicians (Hobbs & Williamson, 2000) found that 47% of respondents reported having opted to perform a maintenance procedure in a way they felt was superior to that described by the manual.

Maintenance technicians also cite problems with unclear or confusing procedures. Sixty percent of respondents in one survey (Hobbs & Williamson, 2000) reported continuation of an unfamiliar task despite not being sure if they were performing it correctly and 67% reported they had been misled by maintenance documentation. Chaparro, Groff, et al. (2002) found that 18% of the respondents reported parts being damaged, 20% reported assembling a component incorrectly, and 25% reported having adjusted or rigged a system incorrectly because of unclear or misleading procedures.

Maintenance Manual Regulatory Requirements

Federal Aviation Regulation (FAR) Part 25.1529 (FAA 2008a) outlines the obligation of manufacturers to provide the technical instructions necessary to support continued airworthiness of the aircraft. The manuals must include information about all equipment installed on the aircraft, including equipment made by third party manufacturers. Manual content requirements are also outlined for system descriptions, maintenance and inspection procedures, required scheduled maintenance, and information about system tests and service points.

The organization of the maintenance manual is specified by the Air Transport Association's Information Standards for Aviation Maintenance (Air Transport Association, 2008) which defines the organizational structure of the AMM and the subject matter to be covered in each chapter. For instance, Chapters 27, 32, and 71 only contain information related to Flight Controls, Landing Gear, and Powerplants, respectively. While the Air Transport Association (ATA) format specifies a high level of organization of the AMM, the formatting, content and level of detail found the chapters differs amongst the manufacturers. Although the FAA requires the manufacturers to provide maintenance manuals, precise requirements regarding the content of the manuals are not defined. The manuals must be accepted by the FAA as part of the aircraft's maintenance program, but the procedural content within the manual itself is not approved by the FAA.

According to FAA regulation FAR § 43.13 (FAA 2008b), an AMT is required to follow procedures outlined in the aircraft maintenance manual. However, there are occasions where situational factors may conspire against strict adherence to the AMM. There can be considerable pressure on AMTs to minimize aircraft down time and return it to service (Hobbs & Williamson, 2000). Under these circumstances, mechanics may be more prone to *workaround* an inadequate procedure rather than contact a manufacturer's technical support for clarification of the maintenance procedure. Unlike maintenance errors that results from the incorrect execution or interpretation of a maintenance procedure the term *workaround* refers to situations where a mechanic is *aware* of a problem with an existing maintenance procedure and then relies on their knowledge and experience or that of their coworkers to identify a means of accomplishing the task. Using the terminology of error analysis the *workaround* is a *violation* because it represents a deviation from standard safe operating practices (Reason & Hobbs, 2003) but unlike other types of violations, they are a response to perceived problems with maintenance documentation.

Research Purpose

The majority of corrections to maintenance manuals are made after their publication (A Chaparro & Groff, 2001). Typically, errors are identified by an AMT per-

forming the procedures who may either contact a manufacturer's customer service engineer for correction/clarification or attempt to identify how to perform the procedure. In the former case, the customer service engineer will verify the error, identify a solution, and decide where to submit a Publication Change Request (PCR) to the technical publications department.

PCRs represent an important source of information regarding the quantity, frequency, and distribution of errors found in the AMM. PCRs can offer insight into why AMTs intentionally deviate from the AMM. PCRs are also likely to be more representative of the errors found in maintenance documentation than analyses based on reports of incidents or accidents. The purpose of this research is to analyze user feedback in the form of PCRs, to document the most frequently reported types of errors found in the AMMs, to identify how errors vary across Air Transport Association (ATA) chapters, and identify the corrective actions taken by the OEM. These data could provide valuable information for the development of interventions to improve maintenance documentation and related causal factors in maintenance errors.

Method

Four aircraft manufacturers, including two general aviation business jet manufacturers and two commercial aircraft manufacturers, agreed to provide PCRs for analysis. Manufacturers were asked to provide a chronological sample of up to 200 PCRs. In one case, this sample represented all of the PCRs pertaining to one aircraft model. As part of the agreement to provide PCRs, which are proprietary documents, the manufacturers were assured anonymity. As such, neither the manufacturers nor the aircraft models are associated with the results. Analysis of this data includes the classification of the types of errors reported, the corrective action requested, and the ATA chapter codes of the requested changes. Only PCRs pertaining to the AMM were included in the analysis.

PCRs were classified using an error taxonomy developed previously (A. Chaparro, Rogers, Hamblin, & Chaparro, 2004). The taxonomy classifies each change request into one of four error types (Technical, Language, Procedural, and Graphics) and 15 error reasons (see Table 2). The associated corrective action made to the manual (i.e., add, delete, or change information) was also recorded for each change request.

Table 2
Error taxonomy used for PCR analysis

	Error Type				
	Technical	*Language*	*Procedure*	*Graphics*	*Effectivity*
Error Reason	Tools	Typos	Step(s)	Part diagram	
	Values/ Tolerances	Grammar/ Terminology	Order	Dimensions	
	Parts	Clarity	Alternative method	Caption/Text	
		Incorrect information	Check/Test/Inspection		
			Caution/Warning		

Four researchers coded the comments contained in the PCRs using the error taxonomy. A Cohen's Kappa (κ) of .78 was calculated on a sample of 25 PCRs reflecting an excellent level of consistency between the coders. Values between .40 and .75 represent fair to good, above .75 as excellent, agreement beyond chance (Fleiss, 1981).

Results

A total of 467 PCRs were analyzed. The PCRs contained multiple change requests and each request for change served as a data point. The result was 879 requests for changes, a mean ratio of 1.88:1 (range 2.7 to 1.6:1) change requests per PCR. Due to the unequal number of PCRs collected from each manufacturer, results will be shown as a percent of total by manufacturer. One manufacturer had sufficient requests related to aircraft effectivity, which warranted the creation of a new category. Effectivity pertains to the applicability of a procedure to a specific airplane. For example, a procedure may be applicable for some aircraft of a given model but not others due to customer modifications or engineering changes implemented in later production aircraft.

Table 3
Breakdown of Error Type for each manufacturer

SOURCE	Error Type (%)				
	Procedural	Language	Technical	Graphic	Effectivity
Manufacturer A	44.5	26.0	14.5	15.0	0.0
Manufacturer B	34.7	28.1	17.4	7.9	12.0
Manufacturer C	42.7	41.2	14.7	1.4	0.0
Manufacturer D	48.2	24.3	19.5	8.0	0.0
M =	42.5	29.9	16.5	8.1	n/a

Table 3 shows the types of errors reported in the PCRs and table 4 shows a summary of the corrective actions (Add, Change, and Delete) requested in the PCRs broken down by manufacturer. Procedural and language errors were the most frequently reported problems; in fact, for all of the manufacturers, procedural and language requests comprise an average 73% of all PCRs (range 62.8% to 83.9%). Almost 90% of the requests involved the change or addition of information to the AMM (see Table 4).

Table 4
Breakdown of corrective actions for each manufacturer

SOURCE	Correction (%)		
	Add	Change	Delete
Manufacturer A	64.6	33.0	2.4
Manufacturer B	48.8	39.3	12.0
Manufacturer C	41.7	43.1	15.2
Manufacturer D	47.8	43.8	8.4
M =	50.7	39.8	9.5

Procedural Errors

The most frequent requests were within the Procedural category. Common procedural errors were categorized as Step(s), Ordering, Alternate method, Check/Test/Inspection, Caution/Warning. Step(s) refers to a request for individual steps within a procedure to be added, changed or deleted; whereas, when a specific type of step(s) was referred to in the PCR, i.e., request for Alternate Method, Check/Test and Caution/Warning, it was recorded. Ordering refers to requests for a change in the sequence of steps by separating, combining, or reordering individual steps.

Table 5
Percentage of corrective actions and types of procedural error requests (PCRs) by manufacturer.

ERROR REASON	Manufacturer	Procedural Errors (%)			
		Corrective Action (%)			
		Add	Delete	Change	% of Total
Step(s)	A	26.0	0	3.0	29.0
	B	20.7	4.2	2.8	27.7
	C	18.0	9.5	2.4	29.9
	D	21.7	5.3	8.8	35.8
	M =	21.6	4.8	4.3	30.6
Order	A	0	0	0	0
	B	0	0	2.3	2.3
	C	0	0	0.5	0.5
	D	0	0	0	0
	M=	0	0	0.7	0.7
Alternative	A	1.0	0	0	1.0
	B	1.9	0	0	1.9
	C	0	0	0	0
	D	0.4	0	0.9	1.3
	M=	0.8	0	0.2	1.1
Check/test	A	7.5	0	0.5	8.0
	B	5.2	0.5	0.5	6.1
	C	4.3	0	1.4	5.7
	D	3.5	0.4	2.7	6.6
	M =	5.1	0.2	1.3	6.6
Caution/Warning	A	5.0	0	0.5	5.5
	B	1.4	0	0	1.4
	C	0.5	0	0	0.5
	D	2.7	0	0	2.7
	M=	2.4	0	0.1	2.5

As shown in Table 5, the most frequently reported Procedural errors were found with the Step(s) category (m = 30.6%). The second category was Check/Test step(s) (m = 6.6%), followed by Caution/Warning step(s) (m = 2.5%).

Language Errors
Language errors found in the PCRs included typographical errors (Typos), grammatical errors (Grammar), a need for clarification of the information (Clarity), and inaccurate information within a step (Incorrect).

Table 6
Percentage of corrective actions and types of Language request reasons via PCRs from each manufacturer.

ERROR REASON	Manufacturer	Language Errors			
		Corrective Action (%)			
		Add	Delete	Change	% of Total
Typo/grammar	A	0	0	0	0
	B	0	0	0	0
	C	0	0	0.5	0.5
	D	0	0	0.9	0.9
	M=	0	0	0.35	0.35
Clarity	A	13.5	1.0	5.0	19.5
	B	15.5	0	2.3	17.8
	C	13.3	1.9	16.1	31.3
	D	9.7	1.3	8.4	19.5
	M =	13.0	1.1	8.0	22.0
Incorrect information	A	0	1.5	5.5	7.0
	B	0	1.9	12.2	14.1
	C	0	3.8	11.8	15.6
	D	0.4	0	4.9	5.3
	M =	0.1	1.8	8.6	10.5

As seen in Table 6, the most frequently reported Language error was Clarity (m = 22%) with users most often requesting additional information (m = 13%) or changing information (m = 8%) to improve clarity. The second most frequent Language error was Incorrect information (m = 10.5%) with users requesting either a change (m = 8.6%) or deletion (m = 1.8%).

Analysis by ATA Chapter
The ATA chapters were also recorded during classification. Analysis shows that the distribution, types of errors reported and requested corrections were similar across the four manufacturers. The most frequent errors reported were found in the chapters related to Flight controls (Chapters 27), Landing gear (32), and Powerplant (71).

Chapter 27 (Flight Controls) was faulted most often followed by Chapter 32 (Landing Gear), both consisting primarily of Procedural and Language errors while most of the errors in Chapter 71 (Powerplant) were Technical. Errors in Graphics were rarely reported. In Chapters 27 and 32, corrective actions to add information were approximately twice as frequent as requests for changes. A small number of requests for deleting information were found in all chapters.

Discussion

We examined the PCRs provided by four aircraft manufacturers to identify the types of errors most commonly reported in AMMs, their distribution across ATA chapters and the types of changes required to address the comments submitted by users. Submission of PCRs was discretionary and consequently may not provide a complete picture of the difficulties experienced by the users. Whiles PCRs do offer insight into the types of errors found in AMMs they provide little insight into the role of organizational culture or situational factors that contribute to errors or the number or types of errors found in other forms of documentation that AMTs use. The study of PCRs complements other types of investigations thus providing a fuller picture of the underlying causes of maintenance error.

Most Commonly Reported Types of PCRs
The results of this study show that the majority of PCRs represent requests for additional procedural information followed by requests to add or change the language to improve clarity. These findings suggest that AMMs may not provide sufficient detail and fail to consider the task from the perspective of the AMT. It is common that an AMT is unable to perform a procedure as described in the manual due to interference from aircraft structures or systems that are not acknowledged in the procedure. Usability testing or proofing by the user population would aid in identification of ambiguous phrasing, poor sequencing of steps, or missing procedural information. Likewise, task analyses would help ensure that procedures exist for tasks commonly performed in the field. AMTs spend much of their time troubleshooting discrepancies on the airplane (e.g., erroneous fuel pressure indicators or inability to control cabin temperature); however, manuals may not include procedures for the sorts of problems that commonly occur during regular aircraft operations (Chaparro et al., 2004).

Errors by ATA Chapter
The majority of PCRs cited procedures found in ATA Chapters 27 (Flight controls), 32 (Landing gear), and 71 (Powerplant). It is interesting to note that the NTSB report of the Air Midwest accident cited procedures in Chapter 27. The rate of occurrence of errors in these chapters may be due to several factors including: 1) a potential reporting bias due to the safety implications of errors in these procedures; 2) the larger number of individual procedures related to these systems and 3) the overall complexity of these systems. As a case in point, consider chapter 71 (Powerplant), which includes maintenance tasks pertaining to electronic sensors, hydraulic & pneumatics, environmental and fuel systems.

Filtering of PCRs

PCRs report discrepancies in the AMM, which AMTs and customer service engineers believe warrant revision of the manual; however, the frequency and distribution of discrepancies found in this study likely represent a conservative estimate. AMTs with more experience or access to experienced co-workers may workaround known errors. Furthermore, AMTs are not equally likely to submit PCRs. This is corroborated by the findings of Chaparro and Groff (2001) that approximately 50% AMTs reported only occasionally, rarely or never reporting errors in the manual. AMTs cited several reasons for not reporting errors including the lack of feedback from manufacturers regarding submitted PCRs and their observation that errors persisted in the manual even after submitting PCRs. Potential PCRs are further culled by customer service engineers, on-site manufacturer representatives, and systems engineers who decide whether a PCR warrants a change in the manual. PCRs, which identify issues that affect safety and technical errors (e.g. incorrect part numbers, settings, clearances etc), receive the highest priority. This emphasis may explain the paucity of PCRs citing the clarity of procedures, spelling errors or typos.

Relationship between Rule Violations, Incidents, and PCRs

The relationship between discrepancies in the AMM and maintenance error is not a direct causal one. In most cases, the discrepancies found in the maintenance manual prompt requests for clarification that delay completion of maintenance tasks. Inspections of work and post maintenance functional tests reduce the likelihood of a maintenance error going undetected. Nevertheless, as recent maintenance related accidents demonstrate, the safety net can fail as the result of the random collusion of factors including a poorly written maintenance procedure, an inexperienced mechanic, time pressure, and failure of the supervisor to review the work and perform functional tests. Improved AMMs would eliminate a source of problems that contribute to errors and mishaps observed downstream.

Figure 1 illustrates the possible outcomes of the executing of a maintenance procedure by an AMT and summarizes the types of rule-related behaviors that may result and where PCRs and incident reports (e.g., report of incident filed in a database MEDA, HFACS-ME, ASRS) are likely to be generated. This figure was adapted from an earlier figure by Reason and Hobbs (2003). Beginning at the top left of the figure the AMT must first identify whether the AMM contains a maintenance procedure for the task. If a maintenance procedure exists then completion of the task will depend on the degree to which the procedure is complete, clear, and correct. The AMT has two choices should no maintenance procedure exist: they may contact the manufacturers' technical support for assistance thus generating a PCR or attempt to accomplish the task by relying on their expertise or that of their peers. The latter choice has several possible outcomes: the task could be performed correctly (i.e., a correct improvisation); alternatively, the task could be performed incorrectly and is detected during functional tests (i.e., a mistake) and an incident report is filed; or finally the task is performed incorrectly and is not detected until a later date (i.e., latent error).
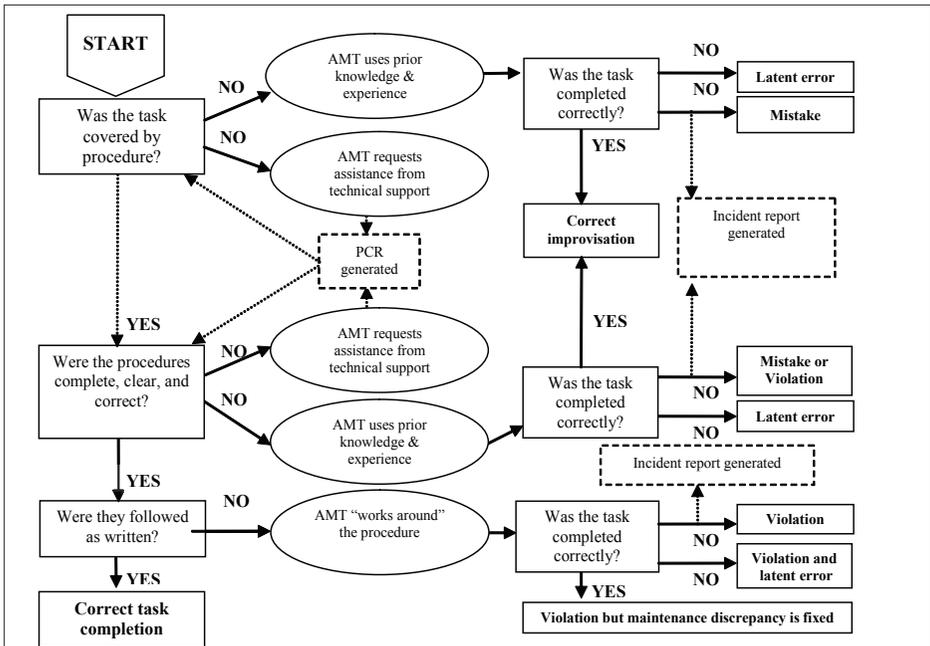
START

Was the task covered by procedure?

NO → AMT uses prior knowledge & experience
NO → AMT requests assistance from technical support

AMT uses prior knowledge & experience → Was the task completed correctly? → NO → Latent error
NO → Mistake
YES → Correct improvisation

PCR generated

Incident report generated

YES

Were the procedures complete, clear, and correct?
NO → AMT requests assistance from technical support
NO → AMT uses prior knowledge & experience

Was the task completed correctly? → NO → Mistake or Violation
NO → Latent error
YES

Incident report generated

YES

Were they followed as written?
NO → AMT "works around" the procedure → Was the task completed correctly? → NO → Violation
NO → Violation and latent error
YES → Violation but maintenance discrepancy is fixed

YES → Correct task completion

Figure 1. Schematic of aircraft maintenance task completion and possible outcomes.

As shown in Figure 1, a maintenance procedure may still fail to meet the needs of the AMT if it is difficult to follow or cannot be executed as described. A maintenance procedure may identify the wrong access panels or fail to take into account physical obstructions that prevent the removal of a component. Again, the AMT may contact technical support or attempt to identify an alternative solution. Finally, there are instances where the manual is technically correct but not followed by the AMT resulting in a *workaround*. *The FAA officially discourages Workarounds* as a maintenance procedure is supposed to be followed exactly as written. This may not always be possible. Technical writers report instances where functional tests described in the manual are technically correct but are not well suited to the task performed by the AMT. One representative example is a case where an AMT was replacing a part of a larger system. The functional test provided in the manual described a complex, time-consuming functional test for the entire system as would occur during assembly of the aircraft. However, no procedure was available for functional testing after the removal and replacement of one subcomponent as is common in the field.

Critical incident reports and PCRs serve several valuable functions. Critical incident error-reporting systems are in place to initiate corrections to the maintenance process within a maintenance facility (i.e., repair station), whereas, the PCR error-reporting system is in place to make corrections to the AMM. PCRs differ from incident reporting systems in the following ways: 1) PCRs are generated at an earlier stage of task completion; 2) PCRs generated by a AMT are fil-

tered at one or more levels within an organization before a change is made to the AMM; 3) the information contained in a PCR is specific allowing for a more fine-grained analysis of problems in the AMM; and 4) PCRs may capture deficiencies in the AMM that may become critical incidents, as well as those that may cause latent errors.

Human Factors and Technical Writing

The problems reported in AMMs are expected given that draft procedures are not evaluated by users; rather, they are reviewed by other writers and engineers. Usually, the technical writer, in consultation with an engineer, writes a draft of a maintenance procedure using engineering drawings, system descriptions, and operational functional tests. The procedure is then circulated amongst members of the technical writing team and system experts for proofing (Chaparro, Rogers, Hamblin, & Chaparro, 2004). However, the emphasis is on technical correctness rather than usability of the procedure. This is evident in the relatively low number of technical errors reported in PCRs. Unfortunately technical correctness does not ensure usability. Some of the problems with maintenance procedures also derive from the fact that they are often based on assembly instructions used on the manufacturing line and consequently do not adequately address typical problems of maintaining an operational aircraft including troubleshooting systems to identify defective components or replacing single components rather then entire systems.

Recent Developments

Following, the Air Midwest accident, the NTSB made two recommendations specifically addressing human factors related issues in maintenance documentation.

Recommendation A-04-13: Require that 14 CFR Part 121 (FAA 2008c) air carriers and aircraft manufacturers review all work card and maintenance manual instructions of critical flight systems and ensure the accuracy and usability of these instruction so that they are appropriate to the level of training of the mechanics performing the work.

Recommendation A-04-16: Require that 14 CFR Part 121 air carriers implement comprehensive human factors programs to reduce the likelihood of human error in (Federal, 2008c) aviation maintenance.

A summary of the FAA response to the NTSB recommendations can be viewed on-line at http://www.ntsb.gov/safetyrecs/private/QueryPage.aspx. In the case of recommendation A-04-13, the FAA responded that the term "critical flight systems" was ambiguous and the FAA would work with the manufacturers to clarify the meaning of the term and develop appropriate procedures. Also, that upon the resolution of another safety recommendation, the FAA would "issue a Fight Standards Information Bulletin (FSIB) to inspectors to ensure air carrier maintenance manuals address the maintenance procedures on critical flight systems." The FAA responded to safety recommendation A-04-16 by stating, "Rulemaking activities will be initiated for 14 CFR 121.375 to require that air carrier maintenance training programs be approved by the FAA."

The NTSB classified both responses as unacceptable noting that in the case of A-04-13 that the "FAA's plan to issue an FSIB does not adequately address the

intent of this recommendation, which is to establish a program to ensure that procedures for flight-critical systems described by airline work cards and maintenance manuals are both accurate and usable." Likewise, the safety board expressed some concern that the FAA did not understand the intent of safety recommendation A-04-16. The board noted that many human factors issues in aviation maintenance including the "availability of proper technical reference documents and guidance, availability of proper and appropriate tools and fixtures, and procedures related to continuation of work from one shift to the next are not related to training" and that a program limited to training would not address these issues. The FAA has not responded to the last round of NTSB comments dated October 12, 2005.

Summary

Analysis of PCRs revealed that the cited problems with the maintenance documentation stem largely from incomplete maintenance procedures and ambiguous phrasing and that these problems are comparable across manufacturers and aircraft size (FAR Part 25 and 121). The problems reported in PCRs are similar to problems cited in previous studies of critical incident reports and surveys.

Previous research has shown that use of user-centered evaluative methods in developing aviation maintenance documentation is effective in revealing potential errors prior to publication of the maintenance manual. These methods may be preferable to the current process that relies on other technical writers, design engineers, and customer technical support engineers in lieu of AMTs. AMTs offer a unique perspective including their familiarity with work place constraints and extensive task-related knowledge that may not be represented by the other groups referenced in the technical documentation development process. This analysis of PCRs reveals similarities in the types of errors reported by users and those found using evaluation techniques including cognitive walkthrough, single-user and co-discovery user performance testing(A. Chaparro et al., 2004).

A breakdown of PCRs as a function of ATA chapters showed that procedures in Chapters 27 (Flight controls), 32 (Landing gear), and 71 (Powerplant) were found to have the highest percentage of PCRs. Given the additional cost of evaluating procedures, this information can be used to develop selection guidelines as to the most critical procedures to evaluate.

Acknowledgements

Author note

Bonnie Lida Rogers is now employed at Cessna Aircraft and Christopher Hamblin is employed at Honeywell Corporation.

References

Air Transport Association of America, Inc. (2008). "ATA Specification 2200 (iSpec 2200): Information Standards for Aviation. Washington, D.C.

Chaparro, A., & Groff, L. S. (2001). *Human Factors survey of Aviation Technical Manuals Phase 1: Manual development procedures* (No. DOT/FAA/AR-01/43). Washington, DC: Federal Aviation Administration.

Chaparro, A., Groff, L. S., Chaparro, B. S., & Scarlett, D. (2002). *Survey of aviation technical manuals Phase 2 report: User evaluations of maintenance documentation*. Washington D.C.: Federal Aviation Administration.

Chaparro, A., Rogers, B. L., Hamblin, C. J., & Chaparro, B. (2004). *Final report: A comparison of two evaluative techniques for validating maintenance documentation*. Washington, DC: Federal Aviation Administration.

Federal Aviation Administration (2008a). Federal aviation regulations. Title 14 of the Code of Federal Regulations 25.1529: Instructions for continued airworthiness, Office of the Federal Register. Washington, DC.

Federal Aviation Administration (2008b). Federal aviation regulations. Title 14 of the Code of Federal Regulations 43.13: Performance rules (general), Office of the Federal Register. Washington, DC.

Federal Aviation Administration (2008c). Federal aviation regulations. Title 14 of the Code of Federal Regulations: 121.567: Certification and Operations: Domestic, Flag, and Supplemental Air Carriers and Commercial Operators of Large Aircraft, Office of the Federal Register. Washington, DC.

Fleiss JL (1981). *Statistical Methods for Rates and Proportions*. John Wiley & Sons, Inc., New York.

Hobbs, A., & Williamson, A. (2000). *Aircraft maintenance safety survey-results*: Department of Transportation and Regional Services, Australian Safety Bureau.

Hobbs, A., & Williamson, A. (2003). Associations between errors and contributing factors in aircraft maintenance. *Human Factors, 45*(2), 186-201.

Lattanzio, D., Patankar, K., & Kanki, B. G. (2008). Procedural error in Maintenance: A review of research and methods. *International Journal of Aviation Psychology, 18*(1), 17-29.

Marx, D. A., & Graeber, R. C. (1994). *Human error in aircraft maintenance*. Aldershot: Avebury Technical: Ashgate.

McDonald, N., Corrigan, S., Daly, C., & Cromie, S. (2000). Safety management systems and safety culture in aircraft maintenance organisations. *Safety Science, 34*, 151-176.

National Transportation Safety Board. (1979). *American Airlines, Inc. DC-10-10, N110AA. Chicago-O'Hare International Airport (Tech. Rep. NO. NTSB/AAR-79-17)*. Washington, DC: Author

National Transportation Safety Board. (2004a). *Brief of Accident (NYC03MA183)*. Washington, DC: Author.

National Transportation Safety Board. (2004b). *Loss of Pitch Control During Takeoff, Air Midwest Flight 5481, Raytheon (Beechcraft) 1900D, N233YV, Charlotte, North Carolina, January 8, 2003* (No. Aircraft Accident Report NTSB/AAR-04/01). Washington, DC: Author.

Patankar, K., & Kanki, B. G. (2001). *Document design strategies for improving airline maintenance procedures.* Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH.

Rankin, W. L., Allen, J. P. J., & Sargent, R. A. (1998). Maintenance Error Decision Aid: Progress report, *The 11th Annual FAA Conference on Aviation Maintenance and Inspection*. Washington, DC.

Reason, J., & Hobbs, A. (2003). *Managing Maintenance Error*. Aldershot, U.K.: Ashgate.

Ricci, K. (2003, October 27-30). *Human factors issues in maintenance publications design.* Paper presented at the DOD Maintenance Symposium & Exhibition, King of Prussia, PA.

Sears, R.L. (1986). A New Look at Accident Contributions and the Implications of Operational Training Programs (unpublished report). Cited in Graeber, R.C. and D.A. Marx (1993). *Reducing Human Error in Aviation Maintenance Operations*. Flight Safety Foundation 46th Annual International Air Safety Seminar and International Federation of Airworthiness 23rd International Conference. Kuala Lumpur, Malaysia. November 8-11, 1993.

Effects of Fatigue on Flight Training:
A Survey of U.S. Part 141 Flight Schools

Sara McDale and Jiao Ma
Department of Aviation Science
Saint Louis University
3450 Lindell Boulevard
St. Louis, MO 63103
USA

Abstract

Fatigue is one human factor that has been found to play a major role in many aviation accidents and incidents. Flight instructors are particularly subject to fatigue due to the traditionally long workday and intensive workload. While there is a plethora of research done on fatigue in relation to commercial pilots and military pilots, little research has concentrated on flight instructors. A national survey was conducted to assess Part 141 flight school instructors' self-awareness of their fatigue issues, impact of fatigue on flight training quality and safety, and potential solutions. Instructors were found to overwhelmingly work long hours, sometimes exceeding 80 hours a week. With the exception of the "too tired to give instruction" scenario, more than half of the instructors reported that fatigue had negatively affected flight instruction in one way or another. Reduced rest and long days were found to be the most common causal factors of instructor fatigue. Survey data raised a concern: flight instructors seem to be unknowledgeable about circadian rhythms and their effect on daily life.

Statistics show that aviation accidents have dramatically decreased since the 1940s, with a leveling off trend since 1978. The improvements in aircraft design, strengthened and redundant systems have removed most of the aircraft and system factors that contributed to accidents. However, the rate of human factor or pilot error accidents (70%) has not changed much since 1940 (Gallagher & DeRemer, 1993). Wiegmann and Shappell (2003) estimate that approximately 70 to 80 percent of all aviation accidents are attributable, at least in part, to some form of human error. General Aviation (GA) accident data for 2006 shows that 73.8% of all accidents and 79.1% of fatal accidents listed pilot error as the major cause (Nall Report, 2007).

Fatigue is one human factor that has been found to play a major role in many accidents and incidents. Michael Mann, an administrator with the National Aero-

nautics and Space Administration (NASA) testified about pilot fatigue in 1999 to the House of Representatives: "it has been evident that pilot fatigue is a significant safety issue in aviation. Rather than simply being a mental state that can be willed away or overcome through motivation or discipline, fatigue is rooted in physiological mechanisms…" (Miller, 2001, p. xviii). Fatigue is described as "the subjectively experienced disinclination to continue performing the task at hand" and that the main effect of fatigue is a "progressive withdrawal of attention from the task at hand" (Miller, 2001, p. 5). Negative effects that fatigue may have on a pilot attempting to operate an aircraft are obvious.

Safety in flight training has always been a major concern due to the number of pilots going through the training process on the path to aviation careers and hobbies. Because flight training is such a significant portion of the aviation industry, this study endeavored to assess the detrimental effects of fatigue in flight training and potential mitigating strategies.

Background

Understanding Fatigue and Pilot Performance

Human causes of accidents/incidents can be attributed to many different factors, including such stressors as fatigue and tiredness that can come from the outside environment, inner emotions, and external interactions. Fatigue is a potentially mentally incapacitating phenomenon that can create a multitude of problems for the flying pilot. Some primary contributing factors for fatigue are lack of quality sleep, stress, health and nutrition, demanding personal schedules and issues, and certain medications (Mayo Clinic, 2008). Salvendy (1987) provides a comprehensive list of subjective complaints related to fatigue. Other researchers have assessed the effects of fatigue on the working memory (i.e., carelessness, forgetfulness, and a reduction in the memory capacity) and judgment (i.e., sloppiness, slow or inappropriate reactions, and loss of timing in performing tasks) (O'Hare, 1999; Jensen, 1995). These symptoms, if not caught and dealt with, can lead to errors and failures by the pilot.

Several factors can attribute to a decrease of pilot performance due to fatigue. Common tasks for a pilot such as flying during the night, elongated cross-country flights, and long hours can result in acute and/or the accumulation of chronic fatigue. Depending on the nature of the flight (personal, professional, business), flights may be conducted after reduced amounts of rest or scheduled without rest in between. As one of the basic human needs, sleep both affects and is affected by numerous lifestyle (e.g., shift work, jet lag, prolonged work hours), socioeconomic, and health related factors (Basner et al., 2007). Sleep restriction induced partial sleep deprivation can cause a range of neurobehavioral deficits, including lapses of attention, slowed working memory, reduced cognitive throughput, depressed mood, and perseveration of thought (Banks & Dinges, 2007). Behavioral alertness is particularly sensitive to sleep loss in general and sleep restriction in particular. According to Colten and Altevogt (2006), performance effects of sleep loss include involuntary microsleeps, unstable attention to intensive performance, increased errors of omission and commission, cognitive slowing in subject-paced tasks, slower response time, deteriorating performance in divergent thinking, and so on.

Other factors that may further increase fatigue include the age of the pilot, the health and fitness of the pilot, and their mental state. In addition, the nature of flight instruction is often repetitious, which can lead to boredom in the training lesson.

Accident data support the fact that many accidents and incidents occur during the takeoff phase and the most occur during the approach and landing phase (FAA, 1999; Nall, 2007). Figure 1 illustrates that pilot capabilities decrease over time, while task requirements mostly increase, causing an unsafe scenario for the fatigued pilot over the course of a flight. Fatigue often plays a major role in committing errors after an extremely long flight or a series of short ones (O'Hare, 1999).
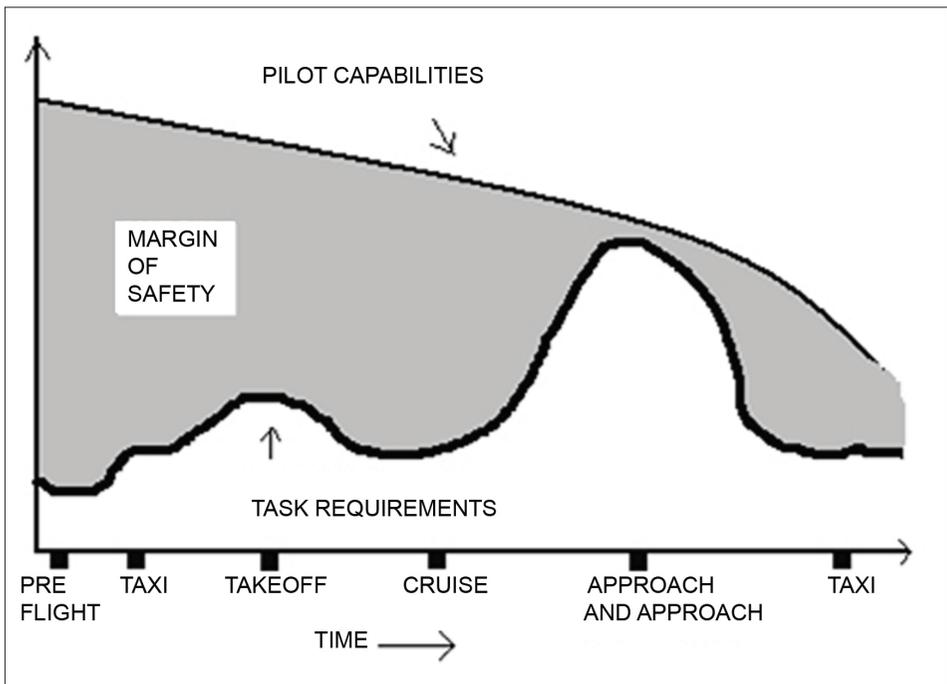
Figure 1. Pilot Performance (FAA, 1999)

Interaction of circadian rhythm ("internal clock" keeps track of the time of day) and sleep-wake rhythm (a process seeks to balance time spent awake and time spent asleep) can potentially influence alertness, fatigue, and performance (Van Dongen & Dinges, 2000). Despite the fact that under normal circumstances the circadian clock continues to exert an influence, a pilot's work may force him/her choose to work or sleep at less than ideal times in the circadian cycle, which induces fatigue and sleep loss. Roehrs and his colleagues found that sleep deprivation produced expected increases in subjective fatigue and impaired performance on measures of attention and cognitive efficiency (Roehrs, Greenwald,

Turner, Koshorek, & Roth, 1999). For example, decision errors are some of the most apparent errors occurring during the nighttime hours for pilots (Wiegmann & Shappell, 2003). Sleep deprivation does not specifically increase impulsive behaviors but decrease risk taking in women but not men (Roehrs et al., 1999; Achesona, Richards, & De Witc, 2007).

Another fatigue-induced contributing factor of accidents could be low levels of arousal. In fact, low arousal can occur at any time, especially shortly after awakening or during extreme fatigue when the nervous system is not fully functioning and the processing of sensory information is slow. Humans' attention mechanisms will not be particularly active at low levels of arousal, and furthermore it is difficult to recover from low states of arousal (Green, Muir, James, Gradwell, & Green, 1996). Due to the circadian biological clock, body temperature and arousal levels vary across a 24-hour period. A circadian trough is a period during the day that a person's body temperature and arousal levels are the lowest (early hours of the morning 3-5 am), thus decreasing the ability to function at a high level of alertness. A second circadian-related dip in arousal is in the afternoon (i.e., "mid-afternoon dip") (Miller, 2001; Van Dongen & Dinges, 2005).

Managing Fatigue
Many people may abide by the attitude of delaying sleep for the rewards of doing something more immediately fun or interesting; however, in aviation, that attitude could spell disasters. Excessive sleepiness may lead to pilot errors, which in the worst-case scenario could lead to death (Miller, 2001).

There are several steps that one can take to either prevent excessive fatigue or to countermeasure existing fatigue. One preventative measure is to minimize sleep loss by taking advantage of one's personal circadian rhythms by allowing sleep when sleepiness is felt (if conditions permit) (Dinges & Broughton, 1989). The Federal Aviation Administration (FAA) released an education and training module on fatigue countermeasures called *Alertness Management in Flight Operations*, which was created in collaboration with NASA's Fatigue Countermeasures Program (Rosekind et al., 1994). Countermeasures to the effects of fatigue while a person is in the cockpit can be very limited. However, there are some ways to keep the body in a state of readiness and alertness, such as small exercises, lively conversation, eating or chewing gum, and drinking moderate amounts of caffeine (Rosekind et al, 1996).

Driskell & Mullen (2005) attempted to integrate the results of 12 studies (a total of 178 separate tests and 270 participant responses) on the effectiveness of naps as a fatigue countermeasure. These studies contained data that allowed precise statistical tests of the effects of naps on performance or fatigue to be derived. Overall, the results of this integration indicated that the average effect of naps for individuals who had been up for an extended period of time was a significant, albeit weak, decrement relative to baseline. The beneficial effects of naps on performance deteriorated after longer postnap intervals. In contrast, fatigue was not affected by nap duration (thus people may report feeling less fatigued after a nap of almost any duration), however, the beneficial effects of naps deteriorated after longer postnap intervals. Circadian rhythms were found not to account for variation in the effects of naps, and no particular association between circadian rhythm and nap duration or postnap interval were found.

The effects of napping on flight crews have been rigorously tested in real-world commercial aviation settings by the NASA and the FAA. After comparing two groups of long-haul flight crewmembers flying the same sequence of four scheduled 9-hour transpacific flights, the crewmembers who were allowed the nap fell asleep on 93% of the available occasions and slept for an average of 26 out of the 40 minutes. They also showed better performance (on a reaction time/ vigilance task) and higher physiological alertness during the last 90 minutes of flight (measured by brainwaves and eye movements) than the control group crewmembers who had not napped (Rosekind et al, 1996). It is apparent that the positive effects of napping can greatly influence the alertness level of pilots, especially during high workload times such as approach and landing.

Most pilots, especially those in the general aviation and training sectors, are familiar with the FAA printed Aeronautical Information Manual's (AIM) recommendation for battling human errors. It is known as the "I'M SAFE" checklist (an acronym for *illness, medication, stress, alcohol, fatigue, and emotions/eating*) (FAA, 2007). The FAA recommends that every pilot examine these seven factors about themselves before flying. If any one of these factors is at an unacceptable level, the flight should be rescheduled or terminated. Many aviators and flight schools abide by this checklist and create a risk analysis before each flight to verify personal airworthiness. A practice such as this can work to increase fatigue awareness and reduce flight error.

General Aviation and Training Accidents

Within the realm of pilot error, fatigue is suspected to be a major component in many unfortunate GA accidents and near accidents (Telfer, 1993; Frazier, 2001; Miller, 2001). A large part of GA is flight instruction, of which these flights are all training flights, usually lasting around an hour to two hours, involving intense maneuvering in short periods of time. A flight instructor, who is legally limited to eight hours of flight in any 24-hour period, can be physically at work for many more hours than that in a given day. For example, several instruction flights may be cancelled due to poor weather, and the instructor would have to schedule multiple students back to back to back on a fair weather day (O'Hare, 1999). Training inexperienced student pilots on those complex flight maneuvers places a high workload on the flight instructor. Previous research has found that fatigue reduces cockpit communication and makes people less rational and more easily irritated (Caldwell, 1997), which affects flight training negatively. In addition, the students are less likely to develop a solid habit on aviation safety and related issues by examples of their fatigued instructors. Because of the above reasons, flight instructor fatigue is potentially more significant than student fatigue, and has a higher detrimental effect on the flight training safety and quality.

The Aviation Safety Reporting System (ASRS) is in place to take proactive measures toward the reduction of incidents and accidents by allowing pilots and others to voluntarily submit information about a flight in which an unsafe situation occurred or report safety hazards. A customized search assisted by the ASRS found 95 GA instruction flights, out of 138,179 reports that that cited fatigue as a

factor. This finding supplies solid data that flight instructors are acutely subject to fatigue related incidents/accidents (ranging from missed frequencies to runway incursions). Considering the fact that all ASRS reports are voluntarily submitted, and potential low awareness and utilization of ASRS reporting by the GA pilots, the aforementioned 95 fatigued instruction flights cannot be considered a random sample of the full population of like events.

Research Aims

Safety of the flight is the ultimate goal of any training or non-training flight. Safety can only be achieved when the crew is alert and aware of all factors concerning the flight: pilot, aircraft, environment, and operation (FAA, 1999). Flight instructors in training flights are particularly subject to fatigue due to the traditionally long workday and intensive workload. While there is a plethora of research done on fatigue in relation to commercial pilots and military pilots, little research has concentrated on flight instructors.

The goals of this study are to evaluate an instructor's self-awareness of their fatigue issues, impact of fatigue on flight training quality and safety, and potential solutions. Most flight instructors are able to identify fatigue issues within the training environment better than students because of higher flight experience and their commitment to ensuring and enhancing safety and quality in flight training. Overall, this study aims at raising awareness of fatigue issues surrounding flight training and increasing the aeronautical decision-making ability of an individual instructor's fitness to fly.

Method

In order to reach a larger number of FAA Certified Flight Instructors (CFIs) effectively, a web-based survey method was adopted. The survey (see Appendix A) is composed of eight sections: 1) work schedule and fatigue level, 2) fatigue and training scenarios, 3) quality of sleep, 4) personal factors contributing to fatigue, 5) ranking of personal factors relating to fatigue, 6) personal solutions to fatigue, 7) ranking of personal solutions to fatigue, and 8) demographic information. Sections 1-3 are designed to understand the instructors' work schedule, general fatigue level, fatigue related to different flight scenarios, and their sleep habits. Sections 4-7 assess actual factors that have contributed to flight instructors' personal fatigue (e.g., schedule, stress, health, etc.) and potential solutions to the fatigue (e.g., reduced workload, more sleep, exercise). The last section collects information about the instructors' age, gender, flight experience, and instruction qualification and experience. The questions were developed based on the related literature and previous survey studies (e.g., Rosekind et al., 2000; Co et al., 1999; Bourgeois-Bougrine, Carbon, Gounelle, Mollard, & Coblentz, 2003).

The survey questions are presented in several different formats such as 7-point Likert rating scales, rank order, and so on. A pilot test of the paper-based survey was conducted with six flight instructors, who represent the typical intended participants, to assess the validity of the questions and usability of the survey (Fowler, 2002). Based on the pilot test feedback, a couple of more questions on quality of sleep were added. Some confusing questions and wording were modified to be more readable. Reliability of the survey was assessed by internal consistency. For

instance, the flight instructors were asked to rank order six time slots on working and non-working days using a "1-most fatigued, 2-slightly fatigued, 3-least fatigued" scale. Consistency check was made to confirm that a given time slot was ranked high on one end and low on the other end of the scale.

Accompanying the survey, a cover letter explained purposes of the survey (see Appendix B), intended respondents, and its importance to aviation safety. CFIs are trained to be very conscious of aviation safety and make every effort to continue their safety education and knowledge. It is logical to assume that most instructors would be very interested in identifying flaws within the flight training system. To foster a trustful relationship with the intended respondents, the cover letter highlighted that the principal investigator (PI) herself is an active CFI.

Participants

Only FAA CFIs were invited to participate in the survey. CFIs are highly trained pilots who have shown competency and proficiency in the areas of instructing, aeronautical knowledge, and various flight maneuvers and techniques. Airplane instructors must hold a commercial certificate with an instrument rating. Only a person who holds a CFI may instruct other people to become certificated pilots. The exception to this rule is that a person with an Airline Transport Pilot (ATP) certificate may only instruct other pilots in the air transportation service (FAA, 2007).

There are two main schools of flight instruction: Part 141 and Part 61. Part 141 programs are FAA approved, generally associated with a collegiate program. They are periodically audited by the FAA and must have detailed, FAA-approved course outlines and meet student pilot performance rates (AOPA, 2008). Thus, Part 141 programs have a higher level of structure and accountability. Because of the quality measures, Part 141 schools are legally allowed to produce certificated pilots with less training time. On the other hand, Part 61 schools are responsible for creating their own training programs and courses and are not subject to routine FAA audits.

In this study, Part 141 instructors were targeted as the main focus. These instructors work with professional aviation students who have the potential to make a large impact on aviation safety as a whole in the future. Part 61 instructors may have some of the same caliber of students, but they also do a lot of work with hobby aviators. In addition, Part 141 instructors are generally easier to contact. In a university setting, computers are easily accessible, thus ensuring that the target population would have the means to partake in a web-based survey.

Procedure

According to 2006 FAA data, there are an estimated 89,452 active flight instructors in the U.S. This data is higher than the targeted population because total flight instructors include glider and helicopter instructors as well as airplane flight instructors (the intended sample). The total number of instructors working in a Part 141 setting is not available, but it would be much smaller than the total

active U.S. instructors. An appropriate sample size was calculated using the sample size formula for the entire flight instructor population (Fink, 2003). An accuracy level of +/- 5% at the 95% confidence level would require 268 surveys be returned. To have an accuracy level of +/- 10% at the 95% confidence level, only 96 respondents were necessary.

The survey was distributed by email to 11 Part 141 flight schools and 1 Part 61 flight school, approximately 175 FAA certified flight instructors. Two reminder emails were sent out. The survey was accessed 84 times between March 1, 2008 and April 8, 2008. Of the 84 survey hits, 74 surveys were completed (response rate of at best 42%). This number is smaller than the required 96 respondents for an accuracy level of +/- 10% at the 95% confidence level to sample a population of 89,452 total active U.S. CFI instructors. However, considering the fact that the intended Part 141 CFI instructors population is much smaller than 89,452, we argue that this sample size is acceptable.

Results

Demographic Information
The average age of the respondents is in the 18-25 years old range, with 64% falling within that category. Only one respondent is over 55 years old. Of all respondents, 82% are males. Ninety-five percent of respondents replied that they are CFIs (a person can still instruct without a CFI as long as he/she holds an ATP certificate). Among them, 22% hold an ATP. Eighty-two hold instrument ratings and 58% hold multi engine ratings on their instructor licenses.

No respondents reported less than 250 hours logged time. About 33% of instructors have logged between 501 and 1,000 hours, and close to 30% logged more than 2,000 hours. This shows that there is a high level of flight experience within the respondent group. Out of 74 instructors, the majority of instructors (71%) have logged at least 251 hours of instruction given time.

Instructors cite experience in many different aspects of instructing, including Part 61 (general pilot training), Part 141 (primary pilot school), Part 142 (pilot school), and Part 121 (air carrier operations). Some respondents (33%) are not currently instructing but may have had recent experience or are familiar with the training environment. Airline pilot hiring started picking up slowly in late 2006. By mid-late 2007, regional airlines were hiring new pilots at low flight time requirements, drawing in a large number of flight instructors (Aviation Schools Online, 2007). This hiring frenzy continued to mid 2008. Many of the survey respondents may have recently moved onto a non-instruction career, thus attributing to the seemingly high percentage of non-active CFIs.

Fatigue and Daily Schedules
When asked if fatigue had a detrimental effect on flight training, the answer is overwhelmingly confirmative (95.9%). Figure 2 illustrates the instructors' typical work schedule in terms of days and hours worked (data of the inactive CFI respondents (n = 24) were excluded). These data may potentially include second jobs because the question did not specify hours worked per day as a flight instructor.
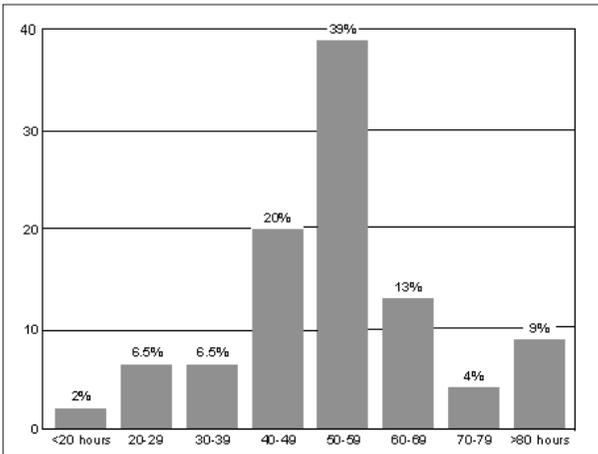
Figure 2. Frequency of instructor hours per week worked by percentage of respondents

When assessing fatigue level on working days (see Figure 3), instructors (n =73) report that they were most tired from 6:00 p.m. to 9:00 p.m. and 3:00 p.m. to 6:00 p.m. The least fatigued time of the day is reported as 9:00 a.m. to noon and noon to 3:00 p.m. On non-working days, the most fatigued part of the day is 6:00 a.m. to 9:00 a.m., followed by the time slot 9:00 p.m. to 6:00 a.m. The least fatigued duration of non-working days is 9 a.m. to noon.
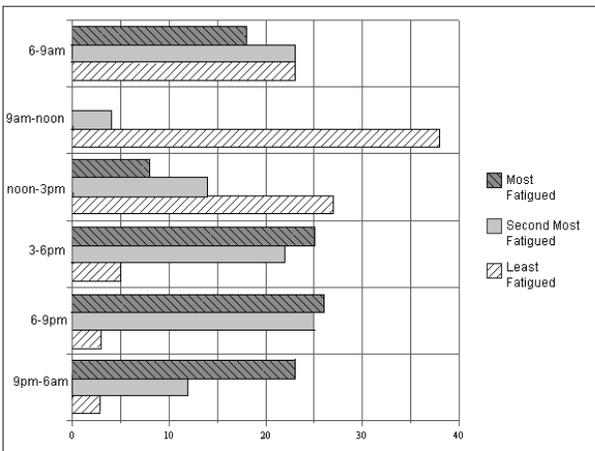
Figure 3. Fatigue levels on a working day by percentage of respondents

Effects of Fatigue

The respondents were queried about their personal experience with the effects of fatigue in flight training. Several scenarios were presented to the instructors and they are asked to either agree or disagree with the statement. Data affirm that many instructors have had experience with fatigue affecting their job (see Figure 4). Most relate with knowing how fatigued they actually are, yet pro-

ceeded to go fly despite of it. Experiencing a disinterest in the flight or irritation toward the student due to fatigue are also common ways that fatigue is manifested.
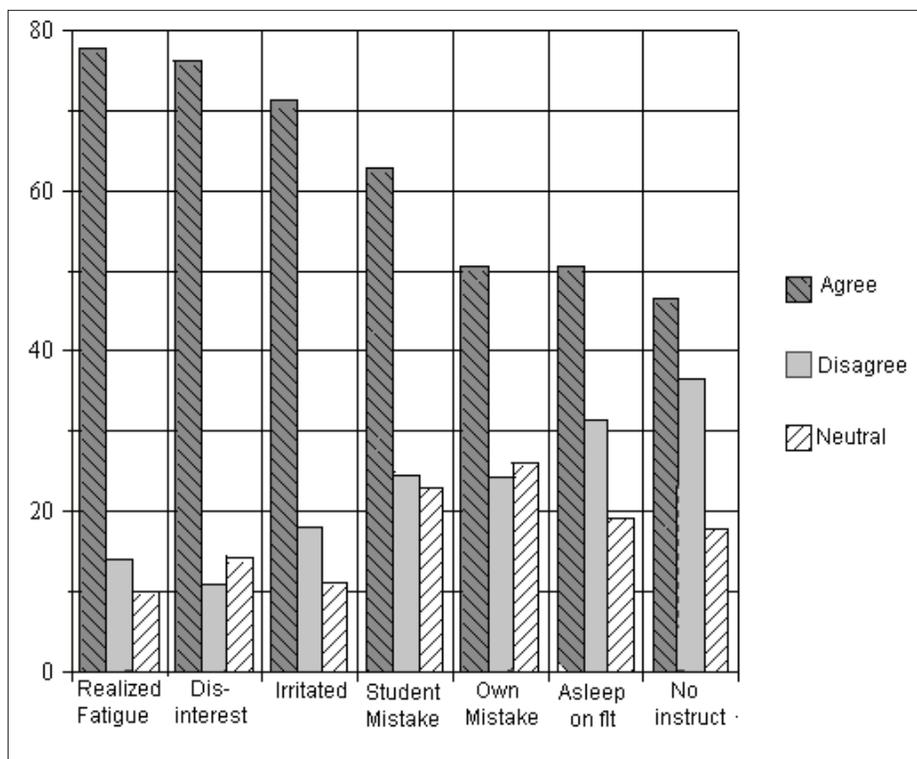

Figure 4. Experiences with the effects of fatigue by percentage of respondents

There is a high level of neutrality on two questions dealing with overlooking mistakes, "To my knowledge, I have overlooked a mistake that a student has/I have made during a training flight." It is likely that many mistakes go unnoticed when fatigue is an issue, thus many instructors might have answered neutral because they had no knowledge of the mistake or reluctant to answer "Yes" (confess of poor training delivered).

Duration and Quality of Sleep
The quality of sleep that an instructor received and its duration could play an important part in the level of fatigue that the person feels. On a regular basis, bedtimes ranged from 9:00 p.m. to later than 2:00 a.m., with the mode as 11:00 p.m. (n = 29). Rise times ranged from 4:30 a.m. to 9:00 a.m., with the mode as 6:00 a.m. The majority of respondents (90.5%) wake up by 7:00 a.m.

Average total sleep received per night is most reported at seven hours (n = 27). The least amount of sleep reported is four hours (n = 1) and the most is eight hours (n = 13). No respondents reported having an average of zero interruptions to their sleep, 62% have one interruption on a typical night. Only 33.8% of instructors claim to feel refreshed after an average night of sleep, while 12.1% percent

respond that they were not refreshed and 54.1% respond "sometimes." Despite conclusive data on how naps can improve performance, especially in aviation, only 13.5% of respondents reply that they generally nap during the day, while 16.2% reply that they sometimes do. There are a number of comments that there is no place to nap at work or that there is not enough time in breaks to accomplish a nap.

Contributing Factors to Fatigue
Instructors were given a list of contributing factors to fatigue to assess which factors had personally affected them in flight training. The results are given in Figure 5.
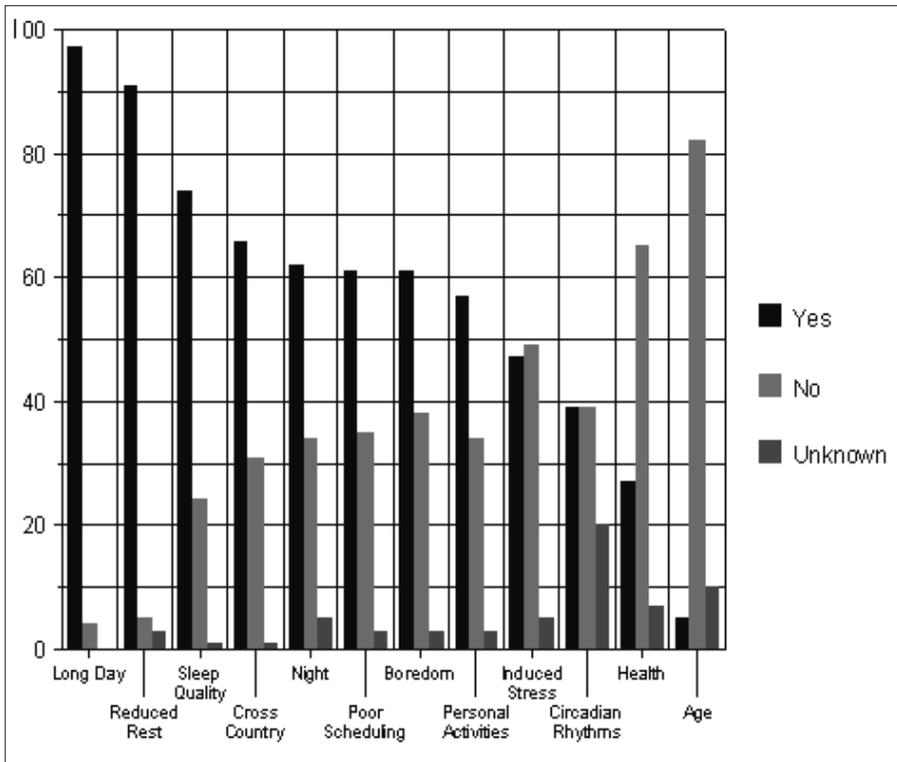

Figure 5. Personal factors contributing to fatigue (n = 74)

The majority of respondents reply that a long day and reduced rest affect their fatigue levels the most. Age and health are found to have the least effect on fatigue. The majority of respondents are under the age of 25, which might explain why many reply that age does not have an effect on fatigue. They may not be aware of the effects due to their comparatively young age. A large number of respondents reply "unknown" when asked if circadian rhythms had an effect on their fatigue level. Many people may not know what circadian rhythms are or how they affect their daily lives. Other factors instructors contribute to fatigue are weather, anxiety and emotion, get-done-itis, not enough days off, and high or dif-

ficult workload. Scheduling and personal reasons are most commented on as contributing factors to fatigue. Several instructors are very passionate in their comments about how schedules were not conducive to getting enough rest.

Respondents are asked to rank the top five factors that had the greatest effect on fatigue (see Table 1).

Table 1
Factors Contributing to Fatigue

Effect on Fatigue	1st	2nd	3rd	4th	5th
Long day	49 %	20 %	16 %	4 %	4 %
Reduced rest	16	23	18	12	8
Boredom	8	10	7	5	14
Scheduling	5	11	7	8	8
Quality of sleep	5	8	10	10	10
Cross country	4	4	5	11	12
Induced stress	3	3	5	12	5
Personal activities	3	8	8	10	10
Circadian rhythms	1	4	4	5	8
Night	0	8	12	14	12
Health	0	0	7	3	3
Age	0	0	0	2	0
Percentage of respondents who answered this question	99 %	99 %	99 %	95 %	93 %

NOTE: all percentages are rounded figures for ease of reading

Solutions to Prevent Fatigue
Instructors were given a list of potential solutions to fatigue as reviewed in the literature. They responded to which solutions had the greatest effect on their personal lives (see Figure 6).
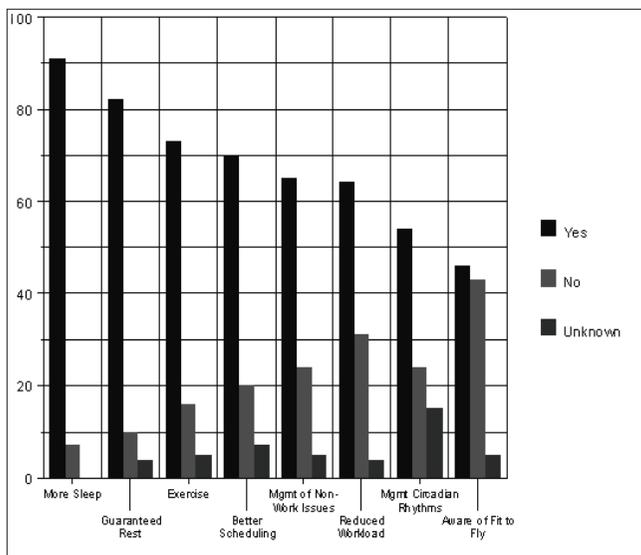

Figure 6. Personal solutions to prevent fatigue (n = 74)

More sleep is chosen as the best solution to prevent fatigue, followed by a guaranteed rest period for a given amount of flying. Respondents do not think that a better self-awareness of their personal fitness to fly would solve any fatigue issues. There again was a relatively high level of "unknown" answers when discussing circadian rhythms.

Respondents propose additional potential solutions to manage instructor fatigue in flight training environment: a place to rest at work, more information or classes on what fatigue is and how to manage it, more pay to decrease the need for a second job, better diet, comfort levels at work and in the plane, improved morale at the workplace. Other comments relate to choices given, such as fitness to fly, scheduling, and more sleep. High workload is not considered a hazard as long as rest was allowed in between flights. Fatigue is something that one respondent feels like he could not talk to his supervisor about because it is unrealistic to say he was "unfit" to fly very often. Another respondent argues that rest in between students is not beneficial unless rest equaled sleep.

Respondents were queried about what they think the best solution to fatigue is by ranking the top five solutions. The best personal solution to fatigue is tied between reduced workload and more sleep, both getting 31% of instructor votes. Guaranteed rest is the next closest best solution with 11% of the votes. These data are summarized in Table 2.

Table 2
Experiences with Personal Solutions to Fatigue

Solution to Prevent Fatigue	1st	2nd	3rd	4th	5th
More sleep	31 %	27 %	10 %	15 %	5 %
Reduced workload	31	12	10	10	10
Guaranteed rest	11	20	19	8	11
Scheduling efficiency	10	12	14	14	10
Fit to fly awareness	5	1	3	4	7
Scheduled breaks	3	11	15	19	10
Management non-work issues	3	4	7	7	15
Management circadian rhythm	3	5	5	11	15
Exercise	1	4	15	8	12
Percentage of respondents who answered this question	99 %	97 %	96 %	96 %	93 %

NOTE: all percentages are rounded figures for ease of reading

Data are analyzed based on the respondents' most popular answers to the effects and solutions to fatigue. If reduced rest is a significant causal factor of fatigue, more sleep seems to be an appropriate solution to the issue. Fifty-nine out of 74 respondents (80%) agree that reduced rest is an issue and more sleep is a potential solution to fatigue. Several instructors (7%) feel that reduced rest affects their fatigue level but do not feel that more sleep would help relieve them of fatigue. Others (5 %) do not think that reduced rest affects their flying, but do feel that more sleep is necessary to combat their fatigue.

Discussion and Conclusions

This research revealed that instructors are aware of the negative effects of fatigue on their performance in flight training. The majority of instructors (71%) have logged at least 251 hours of instruction given time. This shows a high level of experience in flight training. More experienced instructors are supposed to be more familiar with potential fatigue issues in the training environment. Instructors overwhelmingly work long hours, sometimes exceeding 80 hours a week. Reduced rest and long days are found to be the most common causal factors of instructor fatigue. More sleep and reduced workload are the most common solutions proposed to alleviating the fatigue issue.

Knowledge of Circadian Rhythms

Survey data raise a concern: flight instructors seem to be unknowledgeable about circadian rhythms and their effects on daily life. Circadian rhythms are a wired, albeit adaptive, program in the human body. An instructor cannot will him/herself to be alert during all times of the day if it is an innate reaction of the human body to slow down for a few hours. The consequences of not aligning with a person's internal circadian clock may result in a negative influence on his/her alertness, fatigue, and performance levels (Van Dongen & Dinges, 2000).

When assessing fatigue level, the least fatigued time of the day is reported as 9:00 a.m. to noon and noon to 3:00 p.m. Based on extensive research, a person's circadian rhythms slow a body down most during the late night-very early morning (anywhere from 1:00 a.m. to 5:00 a.m.) and a smaller slowing down is also felt around 1:00 p.m. to 3:00 p.m. (Miller, 2001). It is not surprising that the early hours of the morning were not chosen as the most fatiguing time. Flight instruction is largely a day-oriented job and most individuals would be sleeping during the night hours. However, according to the survey, the usual mid-afternoon dip is actually listed as one of the times that instructors felt least fatigued, which indicates that they might not be cognizant of the effects of circadian rhythms, although other factors could be in play. Irregular sleep patterns can also affect how tired one feels during the day.

Awareness of Fitness to Fly

Another trend that merges through data analysis is the fact that most instructors are aware that they work long hard days and subsequently become fatigued. It is their understanding that it is their job to work like that (i.e., good weather days usually equal many back-to-back flights). More importantly, instructors know when they are not at their optimum performance or perfectly fit to fly.

With the exception of the "too tired to give instruction" scenario, more than half of the instructors report that they had a fatigue related occurrence. Seventy-seven percent of instructors have outwardly admitted that they were very tired but proceeded to go fly anyway. These pilots are somewhat aware of their fatigue levels, but are not taking corrective action to regain a solid mental and physical state to fly.

Many reasons contribute to ineffectively attaining solutions to the fatigue problem. Human factors knowledge on fatigue can help raise awareness of potential solutions to the problem and allow pilots to more effectively manage their

fatigue levels. Human factors are often thought as an assumed area of knowledge and are not specifically required for a new certificate or rating (FAA, 2007). The survey data demonstrate an urgent need to keep the knowledge refreshed.

Once the debilitating effects of fatigue are studied and understood, it will be possible to be more cognizant of the breakdowns and mistakes during a flight, thus breaking an error chain. It is common sense for pilots to be aware of the incapacitating results of drinking or taking non-approved drugs while flying. However, many pilots do not stop and think about the consequences of flying while fatigued. Many of the symptoms are the same: lightheadedness and disorientation, inability to collect thoughts or concentrate, dizziness, excessive sleepiness, and reduced reaction time (Salvendy, 1987). Fatigue can be as much a threat to flight training safety as drug or alcohol use. Education and recurrent training on fatigue can have a positive effect on flight training safety.

Limitations of the Study
There are some limitations associated with this survey. The major sources of error in any survey include sampling, coverage, non-response, and what was actually being measured (Gunn, 2002). There is a chance for the exclusion of willing participants who do not have access to the Internet or email due to the distribution method. The flight instructors sampled in this study are close in age to the PI. Of the 89,452 United States active flight instructors, 6,048 of those (6.76%) are female (FAA, 2006), although the percentage is unknown for Part 141 flight instructors. In contrast, 17.6% of the respondents to this survey are females. Because of the PI's professional affiliations and network, the number of sampled instructors may be younger than the national average; and females could be over represented for Part 141 schools. Despite all the effort in ensuring the completion rate (i.e., an informative cover letter, reminder emails, and follow-up contact), non-response error remains a problematic source of survey error (Kent, 2001; Fowler, 2002; Punch, 2003).

The survey intended to collect information on numbers of hours instructors worked per week and the normal timing of their work (e.g., starting time, frequency of early starts) by asking "*what is a typical work schedule for you in any given week? (Give in terms of days a week and hours worked a day, for example, Monday-Friday 9-5).*" However, perhaps because of the wording, respondents only provided typical working hours per week. The timing of flight instructors' work is an important issue to investigate, because sleep and wakefulness that misaligned with internal circadian clock may result in significant sleep loss (Van Dongen & Dinges, 2005).

Future Research
The current study focuses on assessing the awareness of fatigue among flight instructors and impacts on flight training quality and safety. The logical next step would be measure fatigue objectively using equipment such as a sleep watch or some functions of a motionlogger actigraph to document long-term sleep, hyperactivity, daytime activity levels, circadian rhythm, vigilance and so on (AMI,

2008). These data will help quantify the actual sleep received rather than perceived rest. Fatigue can then be more accurately measured. In relation to different levels of fatigue, instruction flights can be assessed in simulated environment. Other than flight performance, cockpit communication, instructors' perceived results of the training session, and student pass rates can be used to evaluate fatigue's impact on flight training.

Best practices and lessons learned from the commercial aviation domain (including fatigue research and initiatives for pilots, mechanics, and air traffic controllers) can be adapted into flight training environment (e.g., Whitehead, 2008; Barimo, Goglia, Nesthus, & Rosekind, 2008). For instance, the Fatigue Risk Management System (FRMS) toolbox for Canadian aviation (Transport Canada, 2008) can be utilized to educate the flight instructors on the basic concept of fatigue and strategies for managing fatigue. Scenario-based training and case studies can be utilized to help instructors develop good judgment skills in their jobs post trainings. Pre-shift and during-shift assessment of sleepiness could be important factors in determining an individual's fitness for duty (Perry, 1998). A fatigue risk management system can be integrated as part of safety management system for Part 141 flight training environment. A fatigue audit system can be used to assess the flight instructors' learning and practice of fatigue management in daily operations.

It is common in Mediterranean, Latin American, and Chinese cultures to take a small mid-afternoon nap during the workweek. Numerous studies have proven the benefits of this resting period in commercial aviation (Milner & Cote, 2008). If more flight instructing schools will implement a napping system for the instructor staff, productivity can greatly increase while fatigue is being staved off, thus reducing the potential for fatigue related error. In a recent review of sleep loss and fatigue in medical personnel, Veasey, Rosen, Barzansky, Rosen, & Owens (2002) offered recommendations for implementing naps and pointed out that naps as short as 15 minutes can significantly ameliorate the performance decrements if provided at 2 to 3 hour intervals. Driskell & Mullen used the regression equation for the effects of naps on performance as a function of nap duration and postnap interval, and consequently provided considerably more precise recommendations for implementing naps (Driskell & Mullen, 2005). The effects of napping in the flight-training environment should be evaluated, since more sleep is considered as one of the best deterrents to fatigue.

A practical and economical solution for North American training schools may be to dedicate a quiet resting area (other than regular office, lounge, or breakroom where instructors meet students, eat meals or perform other activities) for instructors to actually relax and nap.

References

Achesona, A., Richards, J.B., & De Witc, H. (2007). Effects of sleep deprivation on impulsive behaviors in men and women. *Physiology & Behavior*, *91*(5), 579-587.

Ambulatory Monitoring Inc. (2008). Motionlogger. Retrieved May 10, 2008 from the AMI website: http://www.ambulatory-monitoring.com/index.html

AOPA Flight Training. (2008). Choosing your flight school. Retrieved April 16, 2008 from the AOPA website: http://flighttraining.aopa.org/learntofly/school/

Aviation Schools Online (2008). Flight school instructor shortages. Retrieved November 7, 2008 from http://www.aviationschoolsonline.com/

Banks, S., & Dinges, D.F. (2007). Behavioral and Physiological Consequences of Sleep Restriction. *Journal of Clinical Sleep Medicine*, *3*(5), 518-528.

Barimo, B., Goglia, J., Nesthus, T., & Rosekind, M.R. (2008). "The View" of Fatigue Challenges in Aviation Maintenance*. FAA-ATA 20th Annual International Symposium on Human Factors in Maintenance and Ramp Safety*. September 3-4, 2008, Orlando, Florida, USA.

Basner, M., Fomberstein, K.M., Razavi, F.M., Banks, S., William, J.H., Rosa, R.R., et al. (2007). American Time Use Survey: Sleep Time and Its Relationship to Waking Activities. *Sleep, 30*(9), 1085-1095.

Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblentz, A. (2003). Perceived fatigue for short- and long-haul flights: A survey of 739 airline pilots. *Aviation, Space, and Environmental Medicine*, *74* (10), 1072-1077.

Caldwell, J.A. (1997). Fatigue in the aviation environment: an overview of the causes and effects as well as recommended countermeasures. *Aviation Space Environmental Medicine Journal. 68*(10)*,* 931-7.

Co, E. L., Gregory, K. B., Johnson, M. J., Rosekind, M. R. (1999). *Crew Factors in Flight Operations XI: A Survey of Fatigue Factors in Regional Airline Operations* (NASA Technical Memorandum 208799). Moffett Field, California: NASA Ames Research Center.

Colten, H.R., & Altevogt, B.M. (2006). *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. The National Academies Press. Retrieved November 7, 2008, from http://www.nap.edu/catalog.php?record_id=11617

Dinges, D.F. & Broughton, R. J. (Eds.). (1989). *Sleep and Alertness: Chronological, behavioral, and medical aspects of napping*. New York, NY: Raven Press.

Driskell, J.E., & Mullen, B. (2005). The Efficacy of Naps as a Fatigue Countermeasure: A Meta-Analytic Integration. *Human Factors, 47*(2), 360-377.

Federal Aviation Administration. (1999). *Aviation instructor's handbook*. Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (2006). Estimated active pilot certificates held. Retrieved from the FAA website April 2, 2008: http://www.faa.gov/data_statistics/aviation_data_statistics/ civil_airmen_statistics/2006/ media/ air5-06.xls

Federal Aviation Administration. (2007). *FAR/AIM*. Newcastle, WA: Aviation Supplies & Academics, Inc.

Fink, A. (2003). *The survey kit: How to sample in surveys*. Thousand Oaks, CA: Sage Publications.

Fowler, F. J. (2002) *Survey research methods.* Thousand Oaks, CA: Sage Publications.

Frazier, D.A. (2001). *Controlling pilot error: training and instruction*. New York, NY: McGraw-Hill.

Gallagher, R.G. & DeRemer, D. (1993). *Human factors and crew resource management for flight instructors: the new student involvement*. Grand Forks, ND: Eastern Dakota Publishers.

Green, R.G., Muir, H., James, M., Gradwell, D., & Green, R.L. (1996) *Human Factors for Pilots (Second Edition)*. Brookfield, VA: Ashgate.

Gunn, H. (2002). *Web-based surveys: Changing the survey process.* Retrieved April 2, 2008 from the FirstMonday website: http://www.firstmonday.org/issues/issue7_12/gunn/

Jensen, R.S. (1995). *Pilot judgment and crew resource management*. Brookfield, VA: Ashgate.

Kent, R. (2001). *Data construction and data analysis for survey research*. New York, NY: Palgrave.

Mayo Clinic. (2008). *Fatigue: when to rest, when to worry*. Retrieved March 19, 2008 from the Mayo Clinic website: http://www.mayoclinic.com/health/fatigue/HQ00673

Miller, J.C. (2001). *Controlling pilot error: fatigue*. New York, NY: McGraw-Hill.

Milner, C.E., & Cote, K.A. (2008). A dose response investigation of the benefits of napping in healthy young, middle-aged and older adults. *Sleep and Biological Rhythms. 6*, 2-15.

Nall Report. (2007). *Nall Report*. Retrieved March 19, 2008 from the AOPA website: http://www.aopa.org/asf/publications/07nall.pdf

O'Hare, D. (Ed.) (1999). *Human Performance in General Aviation*. Brookfield, VA: Ashgate.

Perry, R.G. (1998). Fitness-for-duty testing. *Occupation Health Safety*, 67, 41-43.

Punch, K.F. (2003). *Survey research: The basics*. Thousand Oaks, CA: Sage Publications.

Roehrs Ta, Greenwald Mk, Turner L, Koshorek G, & Roth T. (1999). Effects of sleep loss on risk taking behavior. *Sleep, 22*(Suppl. 1), 244

Rosekind, M. R., Co, E. L., Johnson, J. M., Smith, R. M., Weldon, K. J., Miller, D. L., Gregory, K. B., Gander, P. H., Lebacqz, J. V. (1994). Alertness Management in Long-Haul Flight Operations. Proceedings of the 39th Annual Corporate Aviation Safety Seminar (pp. 167-178).

Rosekind, M. R., Co, E. L., Gregory, K. B., Miller, D. L. (2000). *Crew Factors in Flight Operations XIII: A Survey of Fatigue Factors in Corporate/Executive Aviation Operations* (NASA Technical Memorandum 2000-209610). Moffett Field, California: NASA Ames Research Center.

Rosekind, M. R., Gander, P. H., Gregory, K. B., Smith, R. M., Miller, D. L., Oyung, R. L., et al. (1996). Managing fatigue in operational settings 1: Physiological considerations and countermeasures. *Behavioral Medicin*e, 21, 157-165. Retrieved March 19, 2008 from http://humanfactors.arc.nasa.gov/zteam/fcp/pubs/j.beh.med1.html

Salvendy, G. (Ed.). (1987). *Handbook of Human Factors*. New York, NY: John Wiley & Sons.

Telfer, R.A. (1993). *Aviation Instruction and Training*. Brookfield, VA: Ashgate.

Transport Canada (2008). *Fatigue Risk Management System for Canadian Aviation - FRMS Toolbox.* Retrieved September 3, 2008 from http://www.tc.gc.ca/tdc/publication/frms/menu.htm

Van Dongen, H.P.& Dinges, D. F. (2000). Circadian rhythms in fatigue, alertness and performance. In: M. H. Kryger, T. Roth & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine* (3rd ed.), pp. 391–399. W. B. Saunders, Philadelphia, Pennsylvania.

Van Dongen, H.P.& Dinges, D. F. (2005). Sleep, circadian rhythms, and psychomotor vigilance. *Clinics in Sports Medicine, 24*(2), 237-49

Veasey, S., Rosen, R., Barzansky, B, Rosen, I, & Owens, J. (2002). Sleep loss and fatigue in residency training: A reappraisal. *Journal of the American Medical Association*, *288*, 1116-1124.

Wiegmann, D.A. & Shappell, S.A. (2003). *A Human Error Approach to Aviation Accident Analysi*s. Bodmin, Cornwall: Ashgate.

Whitehead, B. (2008). Managing fatigue through a systems based approach. *FAA-ATA 20th Annual International Symposium on Human Factors in Maintenance and Ramp Safety*. September 3-4, 2008, Orlando, Florida, USA.

Appendix A:
Effects of Fatigue on Flight Training: A National Survey

(note: survey is shrunk two font sizes to fit into page setup of the appendix)

Section 1 of 8

1.1 Do you feel that fatigue has an effect on the quality of flight training?
      Yes __ No __

1.2 What is a typical work schedule for you in any given week? (Give in terms of days a week and hours worked a day, for example, Monday-Friday 9-5)
_____

1.3 The following are six time slots in a typical flight instruction day. Please rank in order **three** of your fatigue levels from least fatigued to most fatigued.
      1 = you are the **least** fatigued and are still instructing
      2 = you are **slightly more** fatigued and are still instructing
      3 = you are the **most** fatigued and are still instructing

Early Morning (6 am-9am)	1	2	3	N/A
Morning (9am-noon)		1	2	3	N/A
Early Afternoon (noon-3pm)	1	2	3	N/A
Afternoon/Early Evening (3pm-6pm)	1	2	3	N/A
Evening (6pm-9pm)	1	2	3	N/A
Night (9pm-6am)	1	2	3	N/A

1.4 Please rank in order **three** of your fatigue levels from least fatigued to most fatigued on a <u>non-working</u> day.
      1 = you are the **least** fatigued
      2 = you are **slightly more** fatigued
      3 = you are the **most** fatigued

Early Morning (6 am-9am)	1	2	3	N/A
Morning (9am-noon)		1	2	3	N/A
Early Afternoon (noon-3pm)	1	2	3	N/A
Afternoon/Early Evening (3pm-6pm)	1	2	3	N/A
Evening (6pm-9pm)	1	2	3	N/A
Night (9pm-6am)	1	2	3	N/A

Section 2 of 8

Given each scenario, please rank the accuracy of the statement describing your experience as an instructor: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
2.1 I have fallen asleep or struggled to stay awake during a training flight.	1	2	3	4	5
2.2 I have remarked (out loud or to myself) about how tired I was, but proceeded to go on the training flight anyway.	1	2	3	4	5
2.3 I have not given instruction or made a helpful comment to my student because I was too tired to make the effort.	1	2	3	4	5
2.4 To my knowledge, I have overlooked mistakes **a student** has made during the training flight because of reduced awareness or judgment due to fatigue.	1	2	3	4	5
2.5 To my knowledge, I have overlooked mistakes **I** have made during the training flight because of reduced awareness or judgment due to fatigue.	1	2	3	4	5
2.6 I have felt heightened irritation toward my student because I was tired.	1	2	3	4	5
2.7 I have felt a disinterest in the training flight because I was too tired	1	2	3	4	5

Section 3 of 8

Describe your quality of sleep to the best of your ability:
   3.1 Time you usually go to bed on a work day:_____ pm
   3.2 Time you usually rise on a work day:_____ am
   3.3 Number of interruptions to your sleep during an average night: ___

3.4 Do you feel refreshed after an average nights sleep? _____

3.5 How often do you dream? (circle one answer)

     Every night_____Most nights _____ Few nights_____ Rarely_____

3.6 Do you generally nap during the day?
     _____

Section 4 of 8

Based on **your personal** experiences, what factors had contributed to the fatigue you felt during flight training? (Check all that apply and add comments if you desire)

4.1 Flying during night:
     Yes __   No __ Not applicable __ Unknown __
     Comments: _____

4.2 Flying a cross country:
     Yes __   No __ Not applicable __ Unknown __
     Comments: _____

4.3 Working a long day:
     Yes __   No __ Not applicable __ Unknown __
     Comments: _____

4.4 Doing flight training after less than 8 hours of rest:
     Yes __ No __ Not applicable __   Unknown __
     Comments: _____

4.5 Boredom in the lesson:
     Yes __   No __ Not applicable __ Unknown __
     Comments: _____

4.6 Stress caused by family or other psychological conditions:
     Yes __ No __ Not applicable __ Unknown __
     Comments: _____

4.7 Poor scheduling of lessons:
     Yes __   No __ Not applicable __ Unknown __
     Comments: _____

4.8 Circadian rhythms:
     Yes __   No __ Not applicable __ Unknown __
     Comments: _____

4.9 Age:
       Yes __   No __   Not applicable __   Unknown __
       Comments: _____

4.10 Health or fitness:
       Yes __   No __   Not applicable __   Unknown __
       Comments: _____

4.11 Personal activities or other commitments (e.g. 2nd job):
       Yes __ No __   Not applicable __   Unknown __
       Comments: _____

4.12 Quality of sleep:
       Yes __ No __   Not applicable __   Unknown __
       Comments: _____

4.13 Others: _____

Effects of Fatigue on Flight Training: A National Survey

Section 5 of 8

Below is the same list of factors that may have an effect on fatigue. Please read through the entire list and then rank in order five factors from the list based on <u>your personal</u> experiences. (Rank from 1 to 5, with 1 being the greatest effect on fatigue and 5 being the least effect on fatigue)

      _____ **Flying during night**

      _____ **Flying a cross country**

      _____ **Working a long day**

      _____ **Doing flight training after less than 8 hours of rest**

      _____ **Boredom in the lesson**

      _____ **Stress caused by family or other psychological conditions**

      _____ **Poor scheduling of lessons**

      _____ **Circadian rhythms or usual sleep pattern**

      _____ **Age**

      _____ **Health or fitness**

      _____ **Personal activities or other commitments**

      _____ **Quality of sleep**

      _____ **Others**: _____

Section 6 of 8

What do you believe are the best **personal** solutions to prevent fatigue during training? (Check all that apply and add comments if you desire)

6.1 Reduced workload:
Yes __ No __ Not applicable __ Unknown __
Comments: _____

6.2 More sleep:
Yes __ No __ Not applicable __ Unknown __
Comments: _____

6.3 Better efficiency in scheduling:
Yes __ No __ Not applicable __ Unknown __
Comments: _____

6.4 Management of circadian rhythms:
Yes __ No __ Not applicable __ Unknown __
Comments: _____

6.5 Better self-awareness of fitness to fly:
Yes __ No __ Not applicable __ Unknown __
Comments: _____

6.6 Guaranteed rest for a given amount of flying:
Yes __ No __ Not applicable __ Unknown _
Comments: _____

6.7 Physical exercise:
Yes __ No __ Not applicable __ Unknown __
Comments: _____

6.8 Better management of non-work issues:
Yes __ No __ Not applicable __ Unknown __
Comments: _____

6.9 Others: _____

Section 7 of 8

Below is the same list of factors identified as solutions to prevent fatigue. Please read through the entire list and then rank in order five of your <u>personal</u> solutions to prevent fatigue during training. (Rank from 1 to 5, with 1 being the best solution to prevent fatigue and 5 being the worst solution to prevent fatigue)

____ **Reduced workload**

____ **Scheduled breaks**

____ **More sleep**

____ **Better efficiency in scheduling of students**

____ **Management of circadian rhythms or usual sleep patterns**

____ **Better self-awareness of fitness to fly**

____ **Guaranteed rest for a given amount of flying**

____ **Physical exercise**

____ **Better management of non-work issues**

____ **Others**: _____

Effects of Fatigue on Flight Training: A National Survey

Section 8 of 8

Please provide some basic information about yourself so we can better categorize and understand the fatigue issues in flight training. Again, your survey response will remain anonymous. Your organization and the Federal Aviation Administration will not be able to link data and final results from this study to any participating or non-participating individual or organization.

8.1 Age:
__ 18-25   __ 26-35   __ 36-45   __ 46-55
__ over 55

8.2 Gender:
    __ Female   __ Male

8.3 Ratings held (check all that apply):
　　　　 __ CFI 　　　　 __ CFI-Instrument 　　　 __ MEI 　　　　 __ ATP

8.4 Past and current instructing experience (check all that apply):
　　　　 __ Part 61 　　　 __ Part 141 　　 __ Not currently an instructor
　　　　 __ Others, please speci-
fy:_____

8.5 Approximate total logged flight time:
__ less than 250 hours
__ 250 - 500 hours
__ 501 - 1000 hours
__ 1001 - 1500 hours
__ 1501 - 2000 hours
__ more than 2000 hours

8.6 Approximate dual given time:
__ less than 100 hours
__ 101 - 250 hours
__ 251 - 500 hours
__ 501 - 1000 hours
__ 1001 - 1500 hours
__ more than 1500 hours

Appendix B

Survey Cover Letter
Effects of Fatigue on Flight Training: A National Survey

Dear Fellow Pilots:

Hello. I am a graduate student at Saint Louis University seeking my Master's degree in Aviation Safety Management.

I am writing to invite you to participate in this national survey as a part of my ongoing research: *effects of fatigue on flight training*. Your inputs will help en-hance the understanding of how fatigue can affect instruction flights and how to prevent fatigue related accidents and incidents in flight training. Please help, as the data you provide can be crucial in identifying flaws in the flight training system.

This questionnaire is being sent to randomly selected Part 61 and Part 141 flight

schools. Once you have finished answering the questionnaire, please mail/fax/e-mail it directly to

Sara Niemantsverdriet
Center of Aviation Sciences, Saint Louis University
4300 Vector Drive, Cahokia, IL 62206
Tel: 314-977-9580 Fax: 314-977-9541
E-mail: niemansj@slu.edu

The survey will only take about 5 to 10 minutes to complete. Completion and return of this survey indicates voluntary consent to participate in this study. All of your responses are completely anonymous. **There is no way to identify individuals who participate or do not participate in this survey**. The alternative to participation is non-participation. If you decide to not participate in this study, simply discard this survey.
Benefits: There are no explicit or immediate advantages to the subject completing the survey beyond the satisfaction of participating in research aimed at improving aviation safety. **Risks:** The risks associated with participation in this survey are minimal because no identifying data will be collected or held.

Please complete and submit the survey by March 17th. If you choose not to complete the paper copy, an online version is available at:
**http://academic.slu.edu/easysurvey/public/survey.
php?name=FlightTrainingFatigue**

Questions about this study can be directed to me at (314) 977-9580 or niemansj@slu.edu.

Thank you for your input and support of this important study.

Sara Niemantsverdriet, CFI
Graduate Student, Aviation Safety Management
Saint Louis University

Alarm Relevance and Reliability:
Factors Affecting Alarm Responses by Commercial Pilots

Elizabeth T. Newlin
Old Dominion University
Department of Psychology
Norfolk, VA 23529-0267
enewlin@odu.edu
(713)-898-6012

Ernesto A. Bustamante
University of Idaho

James P. Bliss
Old Dominion University

Abstract

The purpose of the present study was to assess the influence of alarm relevance and reliability on pilots' perceptions of alarm relevance, urgency, importance, how compelled they were to respond, and actual response behavior. Twenty commercial pilots flew a simulated aircraft roundtrip from Dulles airport to Boston Logan airport and responded to 20 pressurization alarms. Each participant experienced alarms that were 60% or 80% reliable and alarms with either high or low relevance. Results indicated that pilots performed established flight procedures regardless of system reliability, though they responded faster and more often to more relevant alarms. These findings suggest that pilots consider alarm relevance when responding to alarms but they are extrinsically compelled to respond to unreliable alarms because of their training. The research supports prior alarm prioritization research. Alarm relevance affects pilots' rate and speed of response, and pilots are influenced by their training to overmatch their alarm responses.

False alarms continue to be a pervasive problem in many task domains, including commercial aviation (Bliss, 2003). Although there is extensive research concerning the influence of alarm reliability on pilots' responses (Sorkin, 1988; Bliss, 1997), there is a lack of research concerning influences on pilots' prioritization of alarms. This oversight is problematic because pilots often manage multiple

systems at once. Colvin (2000) and Funk (1991) defined prioritization as the allocation of attention and resources to flight tasks based on relative importance. The prioritization strategy most often advocated by flight trainers is the aviate, navigate, communicate, and systems management (ANCS) hierarchy of task management (Schutte & Trujillo, 1996). The ANCS hierarchy suggests that pilots should perform aviation related tasks before navigation tasks and should prioritize tasks by importance. The ANCS hierarchy provides little information about how pilots prioritize cockpit alarms, although some researchers have suggested that alarm reliability and relevance may influence alarm prioritization strategies (Newlin, Bustamante, Bliss, Spain, & Fallon, 2006). The purpose of the present research is to investigate the joint influence of alarm reliability and relevance on pilots' alarm perceptions and responses.

Alarm Reliability in Aviation

The purpose of an alarm is to notify pilots of a potential problem that requires immediate attention and action (Stanton, 1994a). Alerts, by comparison, indicate potential problems that may require attention in the future (Bliss & Gilson, 1998). Alarm signals frequently include auditory and visual components to capture pilots' attention.

Generally, alarms are designed to activate when the state of the environment exceeds a preset threshold specified by the manufacturer of the alarm system (Parasuraman & Hancock, 1999). Alarms may or may not indicate a genuine problem, depending on the sensitivity of the sensor system and the actual state of the environment (Getty, Swets, Pickett & Gonthier, 1995). The term "Positive Predictive Value" (PPV) refers to the probability that an alarm indicates a true problem. For the purpose of the present study, a low PPV refers to a system that presents a greater number of false alarms. Hereafter the simpler term "reliability" will substitute for PPV.

It is important to further discuss the term "reliability" to avoid confusion. From an engineering point of view, the reliability of a system is typically conceived as the proportion of correct system diagnoses (Wickens & Dixon, 2007). These include both hits or true alarms, and correct rejections. As a result, it is typically an issue of concern to consider false alarms as well as misses when referring to system reliability because they could have differential effects on compliance and reliance (Meyer, 2004). For the purpose of this particular study, we focused on the potential effects of the reliability of a false-alarm prone system.

An alarm system can have low reliability for one of two reasons (Getty et al., 1995). First, there may be a low prior probability that a dangerous event or true problem will occur (Parasuraman & Hancock, 1999). Second, alarm system manufacturers may have set the sensor threshold liberally to detect all possible problems, thereby also generating a greater number of false alarms.

Despite advances in sensor technology and activation algorithms, false alarms are still commonplace. Unfortunately, false alarms cause operators to distrust the alarm system, a phenomenon referred to as the cry-wolf effect (Breznitz, 1984). Lowered trust in turn leads operators to respond less often (Bliss, Gilson & Deaton,

1995; Gupta, Bisantz & Singh, 2002) and more slowly (Getty et al., 1995) to subsequent alarms.

Stanton's (1994b) model of alarm-initiated activities (AIA) suggests how pilots may react to alarms. When an alarm first activates, pilots generally cancel the alarm signal before analyzing the situation. They then gather information from their environment and cockpit displays to verify the existence of a problem. Depending on their analysis, pilots will either ignore the alarm or react to it.

It is essential to emphasize that pilots often make their response decisions in a condition of uncertainty. Because of this, it is possible that they implement one of several strategies, including (a) probability matching, when operators match their alarm response rate to the probability that the alarm represents a true problem, (b) overmatching, which occurs when operators respond to every alarm as if it were true, or (c) undermatching, which refers to ignoring every problem as if it were false (Bliss, 2003). Responding to alarms under uncertainty is considerably more difficult when pilots are inundated with several tasks.

Cockpit Task Management and Prioritization
Prioritization is defined as the allocation of attention based on the importance and urgency of one task compared to other tasks (Colvin, 2000, Funk, 1991). Cockpit task management (CTM) describes how pilots schedule, execute, terminate, shed, monitor, and perform tasks to achieve flight goals (Abbot & Rogers, 1993). Typically, pilots follow the ANCS hierarchy of task management (Schutte & Trujillo, 1996); however, this strategy suggests that each task has an inherent and stable priority. Funk's (1991) theory of task management provides a solid foundation for more recent task prioritization research (Hoover, 2005).

Theories of CTM suggest that pilots schedule tasks based on which ones require their immediate attention. Strategic CTM occurs during normal flight operations, when pilots can anticipate upcoming tasks and plan their prioritization strategy (Rogers, 1996). Tactical CTM occurs during emergency events when pilots must quickly devote available resources to diagnosing and fixing problems. Funk (1991) describes the process of CTM as it occurs in three stages: flight agenda development, analysis of system behaviors, and attention allocation. Funk's (1991) original theory of CTM provides two significant contributions. The theory suggests that prioritization is a cognitive process and that task importance influences task prioritization.

Theories of CTM are useful when describing how pilots manage multiple tasks; yet, to date there has been little prior research concerning alarm prioritization (Newlin et al., 2006). Therefore, research concerning CTM may be useful in understanding the factors that influence pilots' prioritization of alarms. For the purposes of the present study, alarm prioritization is defined as the process of allocating attentional resources to alarms according to their perceived importance, relevance or urgency compared to other attentionally demanding alarms or tasks.

Alarm Relevance

Research concerning CTM suggests that priority determination is situational (Rogers, 1996). In 1996, Rogers added to Funk's (1991) original theory and suggested that CTM is a combination of several processes, including the evaluation of the relevance of tasks to the current situation. For the purpose of the present study, alarm relevance is defined as the connection or association of an alarm to the present circumstances. Therefore, task importance (Funk, 1991) and relevance (Rogers, 1996) are presumed to be imperative for task prioritization; however, pilots likely appraise such factors through a situational filter.

The research conducted by Rogers (1996) and Funk (1991) resulted in a theory of CTM that is applicable to alarm prioritization. Recent research concerning alarm prioritization demonstrated that the relative importance of an alarm influences how operators prioritize multiple alarms (Newlin et al., 2006). Therefore, prior CTM research may provide insight into the factors that influence alarm prioritization.

Flight Research. In 2000, Colvin probed pilots during simulated flights to describe the factors that influenced their task prioritization. Colvin (2000) developed a list of several factors, of which the most influential were task importance and relevance, as predicted by earlier research (Funk, 1991; Rogers, 1996). Colvin's research also lent credibility to situational influences. For example, pilots would prioritize navigation tasks as they neared a waypoint because it was pertinent to the current situation. Schvaneveldt, Beringer and Lamonica (2001) found similar results when they asked pilots to indicate what information was most pertinent during certain flight scenarios. They found that pilots' ratings of importance depended on the relevance of that information to the scenario so that relevant information was regarded as more important.

The distinction between relevance and importance is not trivial. Tasks or alarms may be important but not relevant depending on current operational circumstances. For example, ground proximity alarms are always important because they indicate a potential collision with the ground. However, such alarms are more relevant during takeoff and landing and less relevant at high altitudes and over flat terrain.

Prioritization errors. Prioritization is of particular interest to researchers because misprioritization of tasks often contributes to aviation accidents and incidents. Chou, Madhavan and Funk's (1996) analysis of accident data from the National Transportation Safety Board (NTSB) and the Aviation Safety and Reporting System (ASRS) databases shows that 32% of CTM errors could be attributed to prioritization errors. Although there is little research concerning prioritization errors, the prioritization of alarms has safety implications for commercial aviation.

Goal of this research

Existing research has explored the influence of alarm reliability on piloting behavior, noting the connection between alarm reliability and speed and frequency of alarm responses (Bliss, 1997). However, such research has not considered the concomitant influence of situational alarm relevance. It is conceivable that, faced with uncertain conditions, a pilot may rely on a combination of alarm system reliability and situational relevance to prioritize and react to alarm signals.

Several studies concerning CTM and prioritization (Colvin, 2000; Funk, 1991; Iani & Wickens, 2004; Wickens, Helleberg, Goh, Xu & Horrey, 2001) have suggested that task prioritization demands a more flexible algorithm than the ANCS hierarchy (Schutte & Trujillo, 1996). Many researchers have acknowledged the role of task importance for prioritization (Colvin, 2000; Funk, 1991); yet situational theories suggest that relevance may also be a strong influence (Rogers, 1996; Schvaneveldt, et al., 2001).

The purpose of the present study was to assess the influence of alarm relevance and reliability on pilots' actual response behavior and their perceptions of alarm relevance, urgency, importance, how compelled they felt to respond.

Hypotheses
Perceived relevance hypothesis. Rogers' (1996) theory of task management suggests that operators will continually assess relevance according to situational variables. In this study, pilots were expected to assess the relevance of each PACK high-pressure alarm based on the current altitude of the simulated aircraft. Specifically, pilots were expected to assess PACK high-pressure alarms as more relevant above 25,000 ft MSL, which is the altitude threshold for pressure-related threats to human safety (Federal Aviation Administration [FAA], 2004).

Response rate hypotheses. Participants were required to acknowledge the presence of an alarm by clicking a button. After acknowledging the alarm, participants decided whether they would respond or ignore the alarm and they indicated their choice by clicking a button. Consistent with past research concerning behavioral implications of alarm distrust (Bliss, 1993), the experimenters expected a higher alarm response rate for the more reliable alarm system. Given the results of prior research, pilots were expected to demonstrate a probability matching heuristic (Dorfman, 1969), meaning pilots would respond approximately 80% of the time to the 80% reliable alarm system.

In addition to the main effects for alarm reliability and relevance, the experimenters also expected a significant interaction of system reliability and alarm relevance on response rate. The highest response rate was expected for the more reliable and relevant alarms.

Response time hypotheses. The experimenters expected a significant main effect for alarm system reliability on the time it takes to respond to the alarm. After acknowledging alarms, pilots were expected to respond to alarms from the 80% reliable system faster than those from the 60% reliable system. This hypothesis is based on the influence of alarm reliability on alarm reaction speed (Getty et al., 1995).

Method

Experimental Design
The experimenters employed a within-subject design for this research. The independent variables included alarm relevance and alarm system reliability with

two levels 60% and 80% true alarms. There were two levels for relevance: high at altitudes above 25,000 feet MSL and low below 25,000 feet MSL. The choice of levels for the independent variables was made by consulting with pilots and aviation experts who did not participate in the study. These experts indicated that PACK high-pressure alarms were more relevant at higher altitudes, starting above 25,000 ft MSL, and that the operational reliability of the PACK system approximated 60-80%.

As a primary task, twenty commercial pilots flew a simulated flight using Microsoft Flight Simulator X™ software on two laptop computers. Each pilot completed two flight legs; a flight leg is a segment of a flight from takeoff to landing. The secondary task consisted of acknowledging and responding to PACK (Pressure and Air Conditioning Kit) high-pressure alarms. The PACK high-pressure alarm indicates a potential failure with the PACK device. High-pressure within the PACK stopped the device from cooling and compressing air which may result in lower cabin pressure (Hunt, Reid, Space & Tilton, 1995). Therefore, PACK high-pressure alarms are considered especially relevant above 25,000 ft MSL when adequate cabin pressure is essential for human survivability.

Participants
Twenty commercial airline pilots were selected using snowball sampling. The majority of pilots were male [18 (90%) males and 2 (10%) females]. Pilots were recruited from all airlines; however the pilots in the sample were affiliated with Air Wisconsin, Northwest, and American Airlines. Thirteen (65%) of the pilots were first officers and seven (35%) were captains. Their flight hours ranged from 1,200 to 15,000 (*M* = 3917.87, *SD* = 3354.23) and their ages ranged from 22 to 45 (*M* = 28.23, *SD* = 5.36) years old. All participants had experience with PACK high-pressure alarms and held a commercial pilot certificate with instrument rating or an air transport certificate. The pilots received $50 for their participation.

Procedure
The participants performed a primary flying task on a low fidelity flight simulator. During each flight leg, participants received PACK high pressure alarms. The alarm system was either 60% or 80% reliable and the alarms were more or less relevant depending on the altitude of the aircraft. The experimenters used commercial pilots who did not participate in the study to test all materials, manipulations, and procedures to ensure that they reflected real circumstances during flight.

Each pilot participated individually. To accommodate pilots' busy schedules, the experimenter traveled to a quiet and convenient location for each individual pilot and ensured that ambient room noise did not exceed 50 dB(A). Before participating, pilots read and signed an Informed Consent Form and completed a Background Questionnaire.

Participants received instructions describing the primary flight task as well as the alarm reaction task. In the written instructions pilots read the purpose of the study was to investigate how they would respond to alarms based on the reliability of the alarm and its relevance to the current situation. Pilots were told to consider the altitude of the aircraft when determining the relevance of the alarm. After the experimenter read the instructions, the pilots practiced the flying task and the alarm

response task for five minutes each. Then the pilots practiced the two tasks together for five minutes.

Prior to each flight leg, pilots received information about the reliability of the PACK alarm system in the form of a pre-experimental script and a maintenance log summary. The pilots were told that they could respond to or ignore each alarm based on the information provided about the system's reliability. The experimenters chose the 60% and 80% reliability levels to reflect the range of operational reliabilities found in modern cockpits. The experimental sessions were counterbalanced so that half of the pilots began in the 60% reliability condition and pilots were randomly assigned to the 60% or 80% reliability flight leg.

Primary flying task. Pilots flew a simulated flight using Microsoft Flight Simulator X™ software hosted on an IBM compatible laptop (Windows XP). This software simulated flying a Bombardier CRJ700 aircraft, a 70-passenger regional jet. The Bombardier CRJ700 is a pressurized jet that uses PACK systems to pressurize the aircraft. All pilots were type rated on the CRJ aircraft.

The computer directly in front of the pilots displayed the primary flying task (see Figure 1) on a 12 inch screen and was connected to a joystick that controlled the aircraft. A computer to the right of the pilot displayed simulated air traffic control (ATC) directives on a 14 inch screen.

Figure 1. The pilot's view of the primary flying task.

Pilots flew two 30-min flight legs in clear weather without any turbulence or air traffic from the Dulles Airport in Washington, DC (IAD) to the Logan Airport in Boston, MA (BOS). The experimenters chose these locations based on the need for 30-minute flights, flat terrain, and the need for altitude variability. Pilots began each flight in cruise at 18,000 ft MSL with an indicated airspeed of 250 knots. The pilots were not penalized for heading or speed deviations. They were told to begin each flight as if they had just completed all necessary takeoff procedures. Pilots also ended each flight in cruise at 18,000 ft MSL.

During the flying task, pilots received ATC directives instructing them to ascend or descend to specific altitudes above or below 25,000 ft MSL. A Visual Basic 6.0™ program presented these directives to occur approximately every three minutes. The visual ATC directives were presented on the right hand computer and paired with verbal recordings of an air traffic controller. Pilots spent equal amounts of time above and below 25,000 ft MSL. Pilots were required to manually maintain heading and altitude; this meant they were required to make slight adjustments approximately every 10 minutes to follow the preprogrammed flight plan.

Secondary task. While flying, pilots responded to a total of 20 PACK high-pressure alarms with 10 alarms per flight leg. Alarms consisted of redundant visual and auditory components. A single 600-Hz auditory chime at 60 dB (A) with a duration of 0.5 s represented the PACK high-pressure alarm. These physical parameters match the auditory signal used for PACK high-pressure alarms in the Bombardier CRJ700. A small screen resembling the flight management system (FMS) presented a visual alarm at the same time as the auditory signal. The words "L PACK HI PRESS" or "R PACK HI PRESS" appeared in amber, replicating the format used in the CRJ700.

Alarms were presented according to a variable-interval time schedule with an average of three minutes between alarms so that the pilots were unable to predict their occurrence. Alarms were synchronized to occur at strategic points during each flight leg. When an alarm first occurred, an ACKNOWLEDGE button appeared and the pilots clicked on it while verbally saying, "acknowledge." In accordance with standard commercial airline flight procedures (Quick Reference Handbook (QRH), 2005), pilots were required to acknowledge every alarm. The acknowledge button was replaced by a "RESPOND" button when pilots clicked ACKNOWLEDGE or after 10 seconds (s). Pilots next decided to respond to or ignore the alarm and indicate their choice by either clicking on the RESPOND button or taking no action. The alarm and RESPOND button disappeared when pilots clicked "RESPOND" or after 10 seconds. Setting 10-second time limits for acknowledgement and response was somewhat artificial, but provided an accurate method for measuring response time.

The rest of the alarm procedure mirrored the actual alarm procedure in the cockpit. After pilots clicked on the RESPOND button, they were prompted to complete the PACK high-pressure alarm procedure from the QRH (2005). A section of the control panel appeared on the computer screen and pilots were required to perform the necessary steps in the alarm procedure (see Figure 2). The control panel presented either true or false alarms to match the reliability of the condition (60% or 80%).
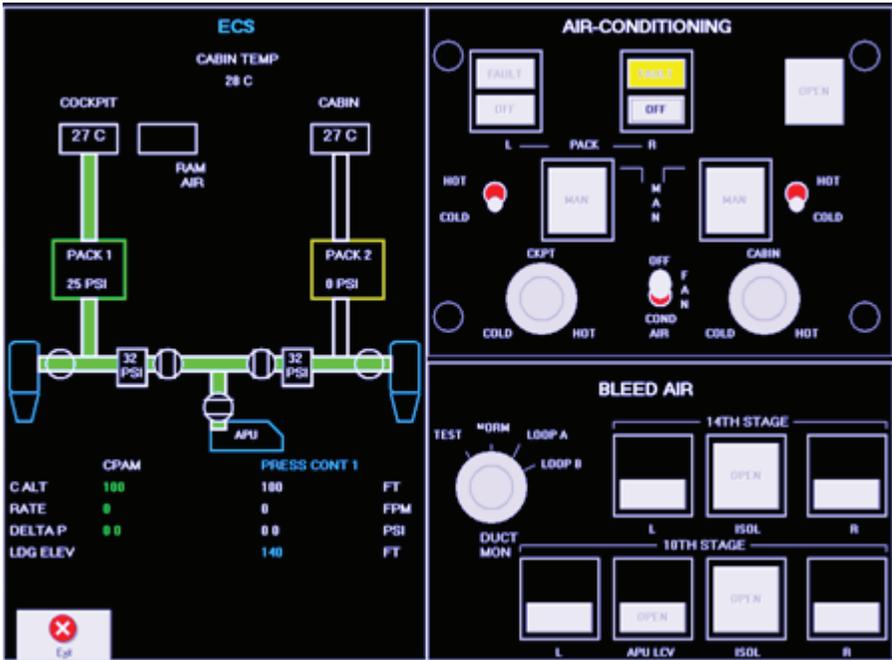
Figure 2. The visual display of the section of the control panel related to the PACK system.

If the alarm was true, the fault light on the air-conditioning panel was yellow and the button below it said OFF. Pilots were required to follow the procedure exactly as it is written in the QRH. The program would not allow them to click the EXIT button until they performed the procedure correctly.

If the alarm was false, pilots would see a normal functioning PACK system as indicated by the absence of a yellow fault light and the presence of green filled lines on the Environmental Control System (ECS) display. In such cases, pilots would click on the EXIT button. Pilots had 30 seconds to complete the alarm procedure before the program recorded their procedure completion time as 30 seconds and the PACK display disappeared automatically.

After the PACK display disappeared, pilots were directed to pause the simulated flight. Pilots pressed the P key on the keyboard to pause the flying task and then answered four post-alarm self-report questions (see Figure 3). The experimenter instructed the participants to consider the altitude of the aircraft and the reliability of the alarm when answering these questions. They indicated how compelled they were to respond to the alarm and the alarm's perceived importance, urgency, and pertinence. The question pertaining to pilots' perceptions of relevance was a manipulation check to ensure that perceived relevance increased as altitude increased. The experimenter instructed participants to interpret relevance as the pertinence of the alarm to the current situation and importance was explained as their perceptions of the necessity to devote attention to the alarm.

Urgency was described as the pilots' perceptions that their response to the alarm could impact the safety of the flight. Pilots indicated their response to each question by selecting a statement and clicking submit. After completing the questionnaire, pilots were prompted to return to the simulated flight by pressing P on the keyboard.

QUESTIONNAIRE

INSTRUCTIONS

Below is a list of statements for evaluating your perception of the relevance of the alarm you just experienced. You should ONLY rate the alarm you experienced. Simply click by the corresponding value that best reflects your opinion about the criteria. You must provide a rating for each and every criterion. Remember your ratings should only reflect your experience with the alarm response you have just completed. Click submit once you have indicated your response.

How compelled did you feel to respond to the alarm?

● 1. I had absolutely no desire to respond at all.

● 2. I did not feel a great desire to respond.

● 3. I felt neither compelled nor disinclined to respond.

● 4. I wanted respond.

● 5. I wanted to respond as quickly and accurately as possible.

Submit

Figure 3. The visual display of the post-alarm questionnaire.

Pilots continued performing the primary and secondary flying tasks until they responded to all 10 alarms programmed to occur during each flight leg. At this time, the Visual Basic 6.0™ program prompted pilots to end the simulation.

For the return flight, if pilots were originally told the alarm system was 60% reliable, they began the second flight leg with an 80% reliable system and vice versa. After the pilot had completed both flight legs, the experimenter conducted a post-experiment interview. The experimenter asked each pilot how the reliability of the alarm system and the altitude of the aircraft influenced their alarm responses. Pilots were also asked if the altitude of the plane affected their perception of alarm relevance. The experimenter also asked if pilots considered responding to alarms as another flight task and what factors might influence their prioritization of multiple alarms or tasks. Finally, the experimenter orally debriefed the pilots.

Measures
Response time was measured in seconds for the amount of time it took pilots to respond to alarms. Response time began once pilots acknowledged alarms by clicking on an ACKNOWLEDGE alarm button, and response time ended either

when they clicked on the RESPOND button or at 10 seconds, representing a maximal response time. Participants who ignored alarms received the maximum response time. Pilots were required to manually fly the aircraft while responding to alarms.

Response rate was calculated as a proportion of responses to the alarm across flight leg. The experimenters chose to measure the proportion of responses rather than the number of responses because proportion data are based on percentages. Calculating proportion data allows easier comparisons to be made with existing and subsequent alarm research. The maximum response rate was 1 which indicated that participants chose to respond to every alarm.

Results

Descriptive Statistics

The experimenters screened all data for outliers, unequal sample sizes, and missing data to ensure normality before completing any statistical analyses. The descriptive statistics showed that response rate, response time, and perceived pertinence were skewed and leptokurtotic; however, ANOVA was robust to this violation of normality because the residuals of the dependent measures were normally distributed (Tabachnick & Fidell, 2001). All data that were 3 standard deviations above or below the mean were considered statistical outliers (Tabachnick & Fidell, 2001). There were no missing data, but there were 2 outliers for response time. These outliers were replaced with the value that was 1 unit larger than the next most extreme value in the distribution, as recommended by Tabachnick and Fidell (2001).

Post-Alarm Questionnaire

Reliability of self-report measures from the questionnaire. Given the repeated-measures nature of the data collection protocol, we examined the reliability of each self-report measure by calculating the average of a series of test-retest reliability estimates. Results indicated that all four self-report measures had acceptable levels of reliability, ranging from .79 to .85.

Correlations among self-report measures. We examined the correlation among all four self-report measures to determine if a multivariate test was required to maintain a family-wide alpha level of .05. Results showed a strong pattern of correlations among all four self-report measures, ranging from .77 to .90.

Multivariate statistical analysis. Given the strong pattern of correlations among the self report measures, we decided to conduct a multivariate analysis of variance (MANOVA) to assess the overall effect of reliability and relevance on the combination of all four measures. Results from a 2 x 2 repeated-measures MANOVA showed a statistically significant main effect for relevance, $F(4,16) = 41.41$, $p < .001$, partial $\eta^2 = .91$.

Univariate statistical analysis. We also conducted univariate analyses of variance (ANOVAs) to further explore the nature of the multivariate effect on each

individual measure. Data from the pilot's answers to post-alarm questions were analyzed by averaging the ratings for alarms that occurred above 25,000 ft MSL and those that occurred below 25,000 ft MSL in both the 60% and 80% reliability conditions. The results of a 2 (system reliability) x 2 (alarm relevance) repeated measures ANOVA showed that pilots were more compelled to respond to alarms when they occurred above 25,000 ft MSL (M = 4.44, SD = .41) than when they occurred below 25,000 ft MSL (M = 3.46, SD = .84), $F(1, 19)$ = 70.31, p < . 001, partial η^2 = .79 (see Table 1). Pilots also indicated that they were more compelled to respond to alarms when they were more reliable (80% reliability; M = 4.07, SD = .75) than when they less reliable (60% reliability; M = 3.82, SD = .88), $F(1, 19)$ = 4.61, p < . 05, partial η^2 = .20 (see Table 2). However there was no significant inter-action of alarm reliability and relevance on motivation to respond (p > .05).

Perceived importance. Results of a 2 (system reliability) x 2 (alarm relevance) repeated measures ANOVA showed that pilots perceived alarms to be significantly more important when they occurred above 25,000 ft MSL (M = 4.32, SD = .45) than when they occurred below 25,000 ft MSL (M = 3.14, SD = .71), $F(1, 19)$ = 131.23, p < . 001, partial η^2 = .87 (see Table 1). There was no significant main effect for alarm reliability and no significant interaction for alarm reliability and relevance on pilots' perception of alarm importance (p > .05).

Perceived urgency. Results of a 2 (system reliability) x 2 (alarm relevance) repeated measures ANOVA showed that pilots perceived alarms to be more urgent when they occurred above 25,000 ft MSL (M = 4.24, SD = .56) than when they occurred below 25,000 ft MSL (M = 2.96, SD = .67), $F(1, 19)$ = 109.59, p < . 001, partial η^2 = .85; see Table 1). There was no significant main effect of alarm reliability and no significant interaction for alarm reliability and relevance on pilots' percep-tion of alarm urgency (p > .05).

Perceived pertinence. Results of a 2 (system reliability) x 2 (alarm relevance) repeated measures ANOVA showed that pilots perceived alarms to be more perti-nent when they occurred above 25,000 ft MSL (M = 4.64, SD = .41) than when they occurred below 25,000 ft MSL (M = 3.12, SD = .81), $F(1, 19)$ = 109.68, p < . 001, partial η^2 = .85; see Table 1). The main effect for alarm reliability and the interaction of reliability and relevance failed to reach statistical significance (p > .05).

Table 1
Means and Standard Deviations for how Compelled Pilots were to Respond and Pilots' Perceptions of Importance, Urgency and Relevance as a Function of Altitude

	Below 25,000 ft MSL		Above 25,000 ft MSL	
	M	SD	M	SD
Compelled to Respond	3.46	.84	4.44	.41
Perceived Importance	3.14	.71	4.32	.45
Perceived Urgency	2.96	.67	4.24	.56
Perceived Relevance	3.12	.81	4.64	.41

Note. The maximum score for these questions was five and the minimum was one.

Table 2
Means and Standard Deviations for how Compelled Pilots were to Respond as a Function of System Reliability

	60% Reliable		80% Reliable	
	M	SD	M	SD
Compelled to Respond	3.82	.88	4.07	.75

Note. The maximum score for these questions was five and the minimum was one.

Hierarchical Linear Models (HLMs)

The experimenters conducted a hierarchical linear model (HLMs) to further explore the potential effects of altitude, system reliability, and the interaction of these two factors on pilots' perceptions of alarm relevance; however, there were no specific hypothesis concerning this test. Reliability was centered at 60%, and altitude was centered at 25,000 ft MSL. The HLM was estimated using restricted maximum likelihood to obtain more adequate and robust estimates of fixed and random effects. Given the fact that there were only two altitude measurement points per participant, random effects associated with altitude could not be estimated. Also, given the repeated-measures of the data and the probability of violating the level-1 homogeneity of variance assumption, statistical significance was set at an alpha level of .01 to reduce the potential of making a type-1 error. Equation 1 represents the fixed (βs) and random effects (rs and e) estimated in the HLM.

$$\text{Dependent Measure} = \beta_{00} + \beta_{10}\text{Reliability} + \beta_{20}\text{Altitude} + \beta_{30}\text{Reliability x Altitude} + r0 + r_1\text{Reliability} + e \qquad (1)$$

The test of the level-1 homogeneity of variance was statistically significant ($p < .01$), suggesting that the homogeneity of level-1 variance was violated.

Pilots' Perceived Alarm Relevance. Results demonstrated a statistically significant main effect of altitude on pilots' perception of alarm relevance, $t(396) = 8.34$, $p < 01$. Pilots' estimated perception of alarm relevance at 25,000 ft was 3.94 s ($\beta_{00} = 3.94$, $SE = .11$) and it significantly increased as the altitude increased in steps of 1,000 ft ($\beta_{20} = .14$, $SE = .02$). No other fixed effects were statistically significant. Figure 4 displays observed and estimated perceived alarm pertinence as a function of altitude.
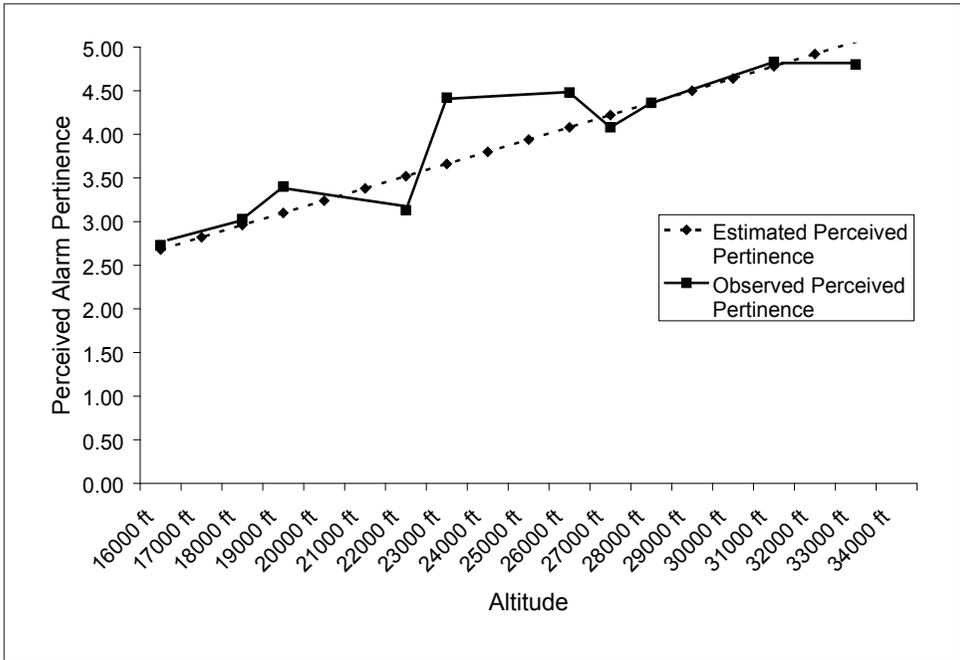
Figure 4. Observed and estimated perceived alarm pertinence as a function of altitude.

Response Rate

Results of the 2 (system reliability) x 2 (alarm relevance) repeated measures ANOVA showed that pilots responded to alarms significantly more often when they occurred above 25,000 ft MSL (*M* = .98, *SD* = .08) than when they occurred below 25,000 ft MSL (*M* = .86, *SD* = .25); *F*(1, 19) = 9.91, *p* < . 01, partial η^2= .34; see Table 3. The proportions for the response rate approached 1.00, indicating that overall pilots responded to 98% of the high-altitude alarms and 86% of the low-altitude alarms. The main effect for alarm reliability and the interaction between reliability and relevance failed to reach statistical significance (*p* > .05).

Table 3
Means and Standard Deviations for Response Rate as a Function of Altitude

	Below 25,000 ft MSL		Above 25,000 ft MSL	
	M	SD	M	SD
Response Rate	.98	.08	.86	.25

Note. The maximum proportion for response rate is one and the minimum is zero.

Response Time

Response time was measured in seconds to assess the time it took pilots to respond to alarms. Results of the 2 (system reliability) x 2 (alarm relevance) repeated measures ANOVA showed that pilots responded faster when alarms occurred above 25,000 ft MSL (*M* = 1.56, *SD* = 1.02) than when they occurred

below 25,000 ft MSL (M = 3.12, SD = 2.41), $F(1, 19)$ = 15.14, p < . 01, partial η^2= .45; see Figure 5). The main effect for alarm reliability and the interaction of reliability and relevance failed to reach statistical significance (p > .05).
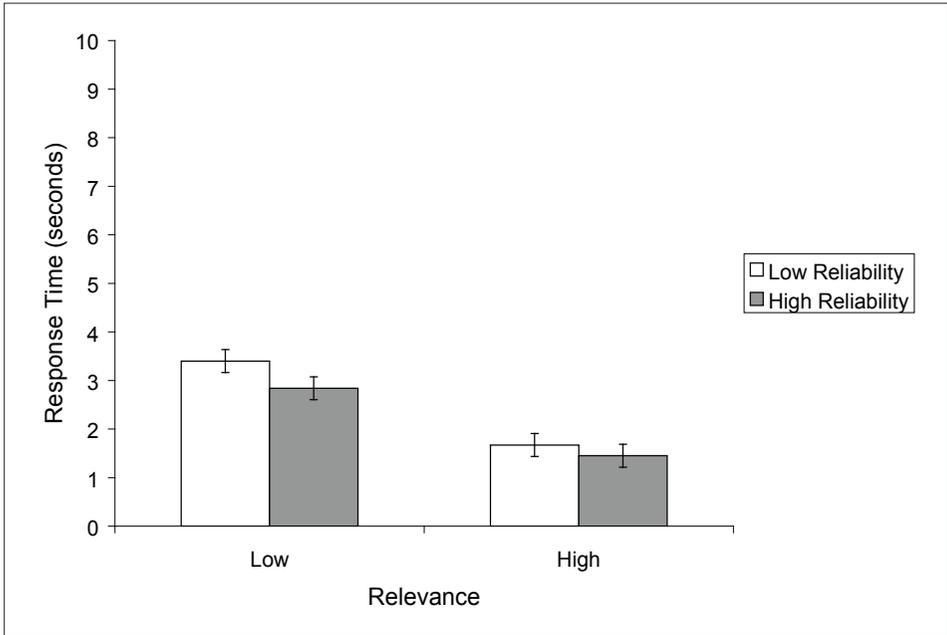

Figure 5. Response time as a function of relevance and reliability.

Discussion

The current research contributes to alarm prioritization literature by investigating the influence of relevance and system reliability on alarm responses.

Theories of Cockpit Task Management (CTM)
Theories of CTM (Colvin, 2000; Funk, 1991; Rogers, 1996) suggest that pilots may perceive PACK high-pressure alarms to be more relevant above 25,000 ft MSL and may therefore respond faster and more often to such alarms. The results fully supported the perceived relevance hypothesis. Pilots considered PACK high-pressure alarms to be more relevant above 25,000 ft MSL than below 25,000 ft MSL and pilots' perceptions of relevance increased as altitude increased. The findings from the post-alarm questionnaire also showed that pilots were more compelled to respond to high-altitude alarms and they perceived those alarms to be more important and urgent.

The results for alarm response rate and response time reflect a similar trend. There was a significant influence of actual alarm relevance on the proportion of times pilots responded to alarms. Pilots responded significantly more often when alarms occurred above 25,000 ft MSL. Furthermore, consistent with the hypothesis regarding response time, results indicated that pilots responded to alarms faster when alarms were more relevant.

These findings suggest that research concerning CTM (Colvin, 2000; Funk, 1991; Rogers, 1996) may be applicable to how pilots determine the priority of alarms. The results demonstrate that pilots may consider relevance when prioritizing between tasks and alarms, as suggested by Rogers (1996). The following quote from a participant, taken during the post-experimental interviews, illustrates the value of relevance:

> First you perform the memory items, and then any other procedures and checklists. So we're trained to prioritize like that. But again it's all in the relevance of the task. If this alarm had happened right before landing at Washington, DC, which is a really dangerous approach and landing, I would have ignored that alarm completely.

The performance findings are consistent with Stanton's (1994b) model of alarm-initiated activities (AIA) that suggests that pilots quickly judge whether to ignore or respond to an alarm, but that they also consider the current situation. Pilots aware of an alarm may fail to take action because the alarm is perceived to be false or interrupts the primary aviating task. In the present experiment, pilots considered the altitude before deciding to respond to the alarm. Once they chose to respond, they accessed the PACK display to investigate and correct the problem by performing the alarm procedure. During the post-experiment interview, several pilots reported that when an alarm occurs they were taught to gather information, assess the situation, and address the most important and relevant tasks or alarms first, rather than reacting immediately. The AIA model (Stanton, 1994b) provides a framework for understanding how pilots consider the context of the situation before responding to alarms.

Alarm Reliability and the Cry-Wolf Effect

Past research concerning alarm reliability and the cry-wolf effect (Breznitz, 1983; Bliss, 1993) demonstrates that complex system operators exhibit greater trust in more reliable alarm systems and subsequently respond more often and faster to signals from such systems. The results from this study failed to support the hypotheses that pilots would respond more often and faster to more reliable alarms. It is important to point out that these unexpected results seem to have occurred mainly because of a ceiling effect. This effect, in turn, could have occurred because of the critical nature of the task at hand. Failing to respond to pressurization alarms can have serious negative ramifications that could place humans' safety in danger. Perhaps this ceiling effect would not be as strong if the cost associated with responding to false alarms were greater than the costs of failing to respond to true alarms. For example, in the case of a weapon deployment task, responding to a false alarm could lead to the wrongful engagement of innocent civilians.

Nevertheless, it is important to emphasize that although alarm system reliability had no influence on pilots' performance measures, pilots indicated that they felt more compelled to respond to the more reliable alarms. These results underscore the variability among response strategies adopted by participants. Prior research (Bliss, 2005) has demonstrated that, whereas most participants tend to match reliability rates with their responses, a certain percentage will elect to

respond to all alarms, overmatch, or no alarms, undermatch. The results of the present study show a tendency to ignore a small percentage of alarms. Some researchers may argue that perfect compliance is the optimal strategy. Yet, the majority of alarm systems are often imperfectly reliable, therefore it may be more effective for pilots to consider how responding to alarms may interfere with their performance of other flight tasks. The findings of the current research suggest that prior training may strongly influence the strategies chosen by participants. Future researchers should explore this possibility to determine the strengths and liabilities of such an influence.

Despite the lack of the statistically significant effect of reliability on pilots' performance, it is important to address these findings from a practical point of view. Based on the fundamental null-hypothesis testing that is typically followed to determine statistical significance, these findings suggest that the reliability of a system has no effect on pilots' performance. However, from a more practical point of view, it is critical to emphasize that our inability to find statistical significance for the effect of reliability is not sufficient to make recommendations to designers and practitioners. From a safety point of view, we do not feel confident in suggesting that a less reliable system is just as effective as a more reliable one, especially given the critical nature associated with the tasks that involve pilots' responses to alarm systems. Our contention is that given specific factors associated with this study, the effect of reliability was perhaps overshadowed by pilots' previous training. Furthermore, as previously mentioned in the introduction, we focused on one aspect of reliability only (i.e., false alarms). Future research is needed to examine if manipulating the reliability of systems by examining false alarms as well as misses would produce different results than the findings of this study.

Commercial Airline Training
Standard commercial airline training conditions pilots to acknowledge every alarm as quickly as possible (QRH, 2005); this may explain the tendency for pilots to respond to most alarms. After the experiment, when asked how pilots incorporated reliability information in their individual alarm responses, 90% of the pilots indicated that they did not consider reliability information. This suggests that many pilots may reflexively respond to alarms before considering their operational relevance or reliability.

During the post-experiment interview, the majority of the pilots indicated that they were extrinsically motivated to respond regardless of system reliability. Anecdotally, pilots felt compelled to investigate each alarm and determine if there was a real problem. Comments indicated that pilots felt uncomfortable ignoring the alarm, even if it was slightly unreliable, because of operational or professional consequences. This perspective may reflect a change between this study and previous research that showed pilots ignore and even deactivate nuisance alarms (Sorkin, 1988). Optimistically, it is possible that commercial aviation trainers have effectively stressed the significance of prompt and consistent alarm reactions.

For some operational systems and circumstances, responding to every alarm is a safe practice. However, there are times when diverting attention toward unreliable alarms represents an unnecessary cognitive burden, particularly during high–workload flight phases such as takeoff and landing. Responding to alarms diverts pilots' attention away from the primary flying tasks. In the current experiment, pilots were aware of the cost of allocating attention to unreliable alarms but chose to respond anyway. When asked why they would allocate their attention to an unreliable alarm, one participant said:

> I guess we always want to do everything even if we have so many things going on. You feel like you should try to juggle all the things that come your way. But I guess it would be better to pick and choose because a lot of times when you take on too much you perform poorly on tasks like flying the plane.

Practical Implications

These findings have significant implications for certain aspects of commercial aviation, especially pilot training. First, the results suggest that theories of CTM apply to alarm prioritization. During the post-experiment interview, the pilots indicated that they considered responding to an alarm to be a task and that they consider alarm relevance, importance, and urgency when prioritizing alarms. Therefore, pilots typically perform the most important and relevant tasks first. Most significantly, the results indicate that relevance is an essential factor in CTM that influences pilots' perceptions of priority and responses to alarms.

Second, prioritization activities performed by pilots may be more complex than the ANCS hierarchy suggests (Schutte & Trujillo, 1996). The results of this experiment indicate that the relevance of a task or alarm depends on the current situation, and pilots consider relevance when responding to alarms. During most emergency situations, pilots prioritize tasks according to the following principles. Pilots are primarily concerned with flying the aircraft; therefore, they focus on maintaining altitude and airspeed. Then, pilots respond to warnings before cautions because warnings are considered higher priority. However, there are circumstances when pilots must consider the situation before responding to alarms. For example, pilots consider the situation of a concurrent left engine failure and a right engine fire alarm. Typically, pilots are trained to respond to an engine fire before an engine failure. In this case, pilots should perform the engine failure procedure first because the engine fire procedure would require an engine shut down, resulting in a crash. This example is reflective of circumstances where the relevance of an alarm depends on the current situation.

These findings have practical applications for pilot training. Hoover (2005) demonstrated that pilots can be taught effective prioritization strategies to reduce CTM errors. Pilots who receive prioritization training commit fewer CTM errors than pilots who receive no training. Hoover (2005) showed reductions in CTM errors when pilots reflected on their prioritization strategies and reviewed these strategies with a flight instructor. In the future, including prioritization-strategy training during commercial-pilot instruction may become increasingly essential because increases in instrument automation and sophistication may lead to CTM errors (Wilson& Funk, 1998). Prioritization training should emphasize alarm relevance, especially when pilots are inundated with multiple tasks and alarms.

Although training has many benefits, it may not be the optimal answer to counter the cry-wolf effect. Training pilots to respond to every alarm places an unnecessary burden on them. For example, pilots may benefit from likelihood alarm displays or systems that alert pilots when to delay responding and when a problem requires their immediate attention (Bustamante, 2005; Bustamante, 2007). These displays would allow pilots to prioritize among tasks and alarms, especially during takeoff and landing. These alarm system redesigns would take the burden off the human and place it on the technology.

Limitations and Future Research

As with all simulation research, the experimental conditions lacked some fidelity. The flight simulator used in the present study was low; the system consisted of two laptop computers. In the future it will be important to investigate the influence of relevance and reliability on pilots' responses to alarms using a higher fidelity flight simulator and a larger sample size. Although commercial aviation pilots typically fly with a co-pilot, the pilots in this study performed alone. The purpose of the present study was to investigate how pilots prioritize alarms depending on alarm relevance and reliability. Including another pilot would complicate the experimental design and make it difficult to solely investigate a single pilot's alarm prioritization strategies. In the future, it will be useful to investigate dyad effects on prioritization strategies, especially since other research using dyads showed an effect for reliability (Bustamante, Fallon, Bliss, Bailey & Anderson, 2005). Further research may show an effect of reliability and relevance on alarm responses when a pilot is part of a dyad.

Another limitation to this study was that pilots did not have access to raw data when an alarm activated. In the future it will be useful to investigate how pilots integrate information from the raw data of a system when deciding to respond to alarms. Further, the present study did not include an evaluation of the impact of alarm relevance or reliability on the pilots' flight performance. In the future, it will be imperative to consider how responding to alarms affects flying performance, especially if researchers wish to consider alarms in the context of CTM.

Future research should focus on expanding the ANCS hierarchy to account for the complexity of task and alarm prioritization. Prioritization is especially complex during compound emergencies when pilots use their systems knowledge to determine how to prioritize alarms. It will be useful to incorporate prioritization literature in the development of a more comprehensive explanation of alarm prioritization. Other scenarios should be developed to study how pilots prioritize concurrent tasks and alarms. It would also be interesting to investigate the influence of time stress and workload on alarm prioritization. These findings will be useful when developing commercial aviation pilot training programs and when designing task support systems (TSS) and other advisory tools in the cockpit.

Conclusion

In conclusion, pilots in this experiment demonstrated that they followed established alarm response procedures despite the reliability of the alarm system. In the future it will be important to consider the influence of reliability on pilots' alarm

response behavior considering that the results contradict prior alarm research. Pilots adhered to their training and considered the relevance of the alarm before responding. Specifically, they perceived the PACK high-pressure alarms to be more relevant above 25,000 ft MSL. This perception caused pilots to respond faster and more often to more relevant alarms. These results support both traditional alarm response research, which suggests that pilots automatically adhere to procedures, and theories of task prioritization, which suggest that pilots consider the relevance of a alarm. Therefore, theories concerning cockpit task management are applicable to the prioritization of alarms. In the future, it will be necessary to further investigate other issues that impact pilots' prioritization of multiple alarms and tasks.

Acknowledgements

References

Abbot, T. S. & Rogers, W. H. (1993). Functional categories for human-centered flight deck design. In *Proceedings of the AIAA/ IEEE 12th Digital Avionics Systems Conference* (pp. 66-74). Fort Worth, TX: AIAA/IEEE.

Bliss, J. P. (1993). *The cry-wolf phenomenon and its effect on operator responses*. Unpublished doctoral dissertation, University of Central Florida, Orlando.

Bliss, J.P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. *The International Journal of Aviation Psychology, 7*(1), 1-14.

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology, 13*(3), 249-268.

Bliss, J. P. & Gilson, R .D. (1998). Emergency signal failure: Implications and recommendations. *Ergonomics, 41*(1), 57-72.

Bliss, J. P., Gilson, R. D. & Deaton, J. E. (1995). Human probability matching behavior in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300-2312.

Breznitz, S. (1984). *Cry-wolf: The psychology of false alarms*. Hillsdale, NJ: Lawrence Erlbaum.

Bustamante, E. A. (2005). A signal detection analysis of the effects of workload, task-critical and likelihood information on human alarm response. In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1513-1517). Santa Monica, CA: Human Factors and Ergonomics Society.

Bustamante, E. A. (2007). Using likelihood alert technology in cockpit displays of traffic information to support free flight. In *Proceedings of the 14th International Symposium on Aviation Psychology* (pp. 100-102). Dayton, OH.

Bustamante, E. A., Fallon, C. K., Bliss, J. P., Bailey, W. R. & Anderson, B. L. (2005). Pilots' workload, situation awareness, and trust during weather events as a function of time pressure, role assignment, pilots' rank, weather display, and weather system. *International Journal of Applied Aviation Studies, 5*(2), 348-367.

Chou, C. D., Madhavan, D. & Funk, K. (1996). Studies of cockpit task management errors. *The International Journal of Aviation Psychology, 6*(4), 307-320.

Colvin, K. (2000). *Factors that affect task prioritization on the flight deck*. Unpublished doctoral dissertation, Oregon State University, Corvallis.

Dorfman, D.D. (1969). Probability matching in signal detection. *Psychonomic Science, 17*, 103.

Federal Aviation Administration. (2004, May). *Supplemental oxygen for emergency descent and for first aid; turbine engine powered airplanes with pressurized cabins*. Retrieved May 19, 2007, from http://a257.g.akamaitech.net/7/257/2422/12feb20041500/edocket.access.gpo.gov/ cfr_2004/jan-qtr/pdf/14cfr121.333.pdf.

Funk, K. H. (1991). Cockpit task management: Preliminary definitions, normative theory, error taxonomy and design recommendations. *The International Journal of Aviation Psychology, 1*(4)*,* 271-285.

Getty, D. J., Swets, J. A., Pickettt, R. M. & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*(1)*,* 19-33.

Gupta, N., Bisantz, A. M. & Singh, T. (2002). The effects of adverse condition warning system characteristics on driver performance: An investigation of alarm signal type and threshold level. *Behaviour & Information Technology, 21*(4)*,* 235-248.

Hoover, A. L. (2005). Experimental analysis of task prioritization training for a group of university flight technology students. Unpublished doctoral dissertation, Oregon State University, Corvallis.

Hunt, E. H., Reid, D. H., Space, D. R. & Tilton, F. E. (1995, May). *Commercial airliner environmental control system: Engineering aspects of cabin air quality.* Paper presented at the meeting of the Aerospace Medical Association, Anaheim, CA.

Iani, C. & Wickens, C. D. (2004). *Factors affecting task management in aviation.* (Report No. AHFD-04-18/NASA-04-7). Moffett Field, CA: NASA Ames Research Center.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*, 196-204.

Newlin, E. T., Bustamante, E. A., Bliss, J. P., Spain, R. D. & Fallon, C. K. (2006). The effects of relative system reliability and prioritization on alarm reaction patterns. In *Proceedings of the 50th Human Factors and Ergonomics Society Annual Meeting* (pp. 1675-1679). San Francisco: The Human Factors and Ergonomics Society.

Parasuraman, R & Hancock, P. A. (1999). Using signal detection theory and Bayesian analysis to design parameters for automated warning systems. In M. W. Scerbo & M. Mouloua (Eds.), *Automation, technology and human performance: Current research and trends* (pp. 63-67). Mahwah, NJ: Erlbaum.

Quick Reference Handbook (QRH). (2005). *CL-65 (CRJ 200): Quick reference handbook.* (Available from the Air Wisconsin Airlines Corporation, W6390 Challenger Drive, Suite 203, Appleton, WI 54914)

Rogers, W. H. (1996). Flight deck task management: A cognitive engineering analysis. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 239-243). Philadelphia, PA: The Human Factors and Ergonomics Society.

Schvaneveldt, R. W., Beringer, D. B. & Lamonica, J. A. (2001). Priority and organization of information accessed by pilots in various phases of flight. *The International Journal of Aviation Psychology, 11*(3)*,* 253-280.

Schutte, P.C. & Trujillo, A. C. (1996). Flight crew task management in non-normal situations. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 244-248). Philadelphia, PA: The Human Factors and Ergonomics Society.

Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of the Acoustical Society of America*, *84*(3), 1107-1118.

Stanton, N. (1994a). A human factors approach. In N. Stanton (Ed.), *Human factors in alarm design* (pp. 1-10)*.* London: Taylor & Francis Ltd.

Stanton, N. (1994b). Alarm-initiated activities. In N. Stanton (Ed.), *Human factors in alarm design* (pp. 93-117)*.* London: Taylor & Francis Ltd.

Tabachnick, B. G. & Fidell, L. S. (2001). *Computer-Assisted Research Design and Analysis.* Boston, MA: Allyn and Bacon.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201-212.

Wickens, C.D., Helleberg, J., Goh J., Xu, X. & Horrey, W. J. (2001). *Pilot task management: Testing an attentional expected value model of visual scanning*. (Report No. ARL-01-14/NASA-01-7). Moffett Field, CA: NASA Ames Research Center.

Wilson, J. R. & Funk, K. (1998). The effect of automation on the frequency of task prioritization errors on commercial aircraft flight decks: An ASRS Incident Report Study. In *Proceedings of the Second Workshop on Human Error, Safety, and System Development* (pp. 6-16). Seattle, WA.

Stress in Ballooning: An Exploratory Cortisol Study

Alex de Voogt
Dept of Work & Social Psychology
Faculty of Psychology
Maastricht University
PO Box 616, 6200 MD
Maastricht, The Netherlands

Abstract

Pilot errors are prominent in aviation and this may partly be mediated by stress-induced hormonal changes that deteriorate cognitive information processing. The present study explores the possible existence of stress in balloon operations by measuring stress-related hormonal changes during balloon flights. Salivary cortisol was measured in experienced balloon-pilots before, during and after a balloon flight. Compared to pre-flight cortisol concentrations, cortisol significantly increased 20 minutes after take-off during flight. The data suggest that even in experienced pilots, balloon flights may be stressful and therefore may influence the risk for pilot errors. Further research is necessary to generalize this claim

Stress in Ballooning: An Exploratory Cortisol Study

The prevalence of accidents in general aviation is generally attributed to pilot error (Li, Baker, Grabowski, Rebok 2001). The attention and concentration necessary to perform well as a pilot are affected by psycho-physiological factors, such as stress and fatigue (Cedhara, Hyde, Gilchirst, Tytherleigh, Plummer 2000). Keeping in mind that cortisol is part of an adaptive response system involved in the mobilization of energy and resources to cope with challenges (Munck, Guyre, Holbrook 1984), cortisol-levels are used as an indicator of stress (Ursin & Olff 1993) and have also been used to study demanding tasks in flight operations.

Demanding tasks have ranged from carrier landings (Miller 1970) and simulated bird-strikes during flight (Sive & Hattingh 1991) to flight tests (Kobayashi 1996) and instrument flights for students (Leino, Leppaluoto, Ruokonen, Kuronen 1999). In these studies students showed higher cortisol levels after such flights than instructors (Kobayashi 1996), and pilots more than radar men (Miller 1970),

for instance. Hot-air ballooning may be seen as an example on the less-demanding side of the spectrum, but as in most aviation operations, balloon operations feature accidents where pilot errors are prominent (de Voogt & van Doorn 2006). In ballooning, take-offs and landings are particularly dangerous since the balloon is at low altitude and the pilot has limited control.

This study was conducted to determine whether stress is a factor with pilots in balloon operations. For this purpose, salivary cortisol concentrations were measured in experienced balloon pilots before, during and after a balloon flight.

Methods

Hot-air balloonists participating in a ballooning event in the Netherlands were approached during a briefing to participate in this study. Nine male pilots volunteered with an age ranging from 35 to 63 years. Their experience in ballooning was between seven and twenty years. Flight experience ranged from 240 to 1600 hours, with four pilots showing at least 1000 hours. Two pilots did not provide data on age and experience.

All flights were conducted late afternoon during the same day and following the same time table, to ensure comparable weather circumstances. On the day of the balloon event, the balloonists were tested for cortisol responses before, during, and after their balloon-flight. Shortly after the weather briefing preceding the flight, each pilot was instructed to obtain a salivary sample before take-off and prior to operating the burner that inflates the balloon with hot air (pre-flight), at approximately 20 minutes after take-off (in-flight), and within 15 minutes after landing (post-flight). All balloonists carried two to four passengers and the total time of flight was approximately 45 minutes for each pilot. After their return to the take-off site all salivary samples were collected and frozen at -25 °C. In addition, information was obtained concerning time of salivary collection, consumed food and drinks before or during the flight, and possible particular circumstances of the landing.

Cortisol samples were obtained by using the Salivette sampling device (Sarstedt®, Etten-Leur, Netherlands). Saliva samples were centrifuged at 2650 g_{max} for three minutes at 20 °C. Salivary cortisol levels were determined by direct radioimmunoassay (RIA; University of Liège, Belgium).

The data were analyzed by means of repeated measure analyses of variance (ANOVA) using the General Linear Model (GLM: SPSS 7.5 for Windows) with one within-subject factors "*Time*" (pre-flight, in-flight, post-flight) on cortisol concentrations as the dependent variable. Statistics were evaluated at a significance level of 5%. Data are reported as means ±SD.

Results

The take-off procedure was conducted in fair weather circumstances and without incident. All landings were considered uneventful. Two pilots reported strong winds with other balloons landing in the vicinity but without experiencing further inconvenience. None of the balloons tipped over during the landing.

No influence from sugar, tobacco or coffee intake is suspected. Only one pilot reported drinking something just prior to the flight (ice tea). All saliva samples were collected within the specified time range, although one pilot forgot to collect the in-flight sample so that for most parts of the analysis only eight samples were used.

All participants that provided an in-flight and pre-flight saliva sample showed an increase in cortisol levels of, on average, 140%. All but one of these participants showed a subsequent decrease of cortisol levels from the in-flight to the post-flight measure.

A first overall repeated measures analysis of variance with *Time* (pre-flight, in-flight, post-flight) as within-subject factors on salivary cortisol concentrations revealed a trend for significance of *Time* [$F_{(2,14)}=4,37$; *P*=0.072]. This trend analysis only approaches significance, but a second analysis on cortisol changes revealed a significant increase in cortisol in-flight (17.74±12 nm/L) as compared to pre-flight (7.07±4 nm/L) [$F_{(1,7)}=6,73$; *P*=0.036], whereas analysis on post-flight relative to in-flight cortisol changes did not reveal any difference (16.42±15 nm/L) [$F_{(1,7)}= 0.011$; $P>0.9$].

Conclusion

The current study examined salivary cortisol responses during a balloon flight in a group of experienced balloon-pilots. Take-off resulted in an average of 140% increase of cortisol concentrations when comparing pre-flight with in-flight, while there were no significant changes in cortisol levels shortly after landing.

For eight balloonists in this study, cortisol levels increased from the pre-flight to the in-flight measure. Since cortisol is seen as the main stress-related hormone released by activation of the pituitary-adrenocortical (HPAC) system (Munck, Guyre, Holbrook 1984), this flight-induced increase in cortisol suggests an increase in stress experience during the balloon flight.

Although the balloon event itself may have increased stress levels due to the crowded space and simultaneous balloon lift-offs, it is noted that most pilots had participated in such events before and that weather circumstances, despite some strong winds during the landing, were fair and not dangerous or particularly demanding.

Ballooning may not be considered a high-risk or a highly demanding part of aviation, but a possible stress increase suggests that this does not eliminate the influence of psycho-physiological factors on the risk for pilot errors. The flight experience of the balloonists in the sample indicates that a possible stress increase during balloon operations is not necessarily associated with students but affects various experience levels.

Further research may indicate whether cortisol levels decrease in long flights, commonly winter flights that may last about two hours, and whether the anticipa-

tion of a landing then leads to a cortisol level increase. Other measures, in particular heart rate variability and larger sample sizes are necessary to generalize the findings in this study.

References

Cedhara K., Hyde J., Gilchirst I.D., Tytherleigh M., & Plummer S. (2000). Acute stress, memory, attention and cortisol. *Psychoneuroendocrinology 25*(6), 535-549.

de Voogt A.J., & van Doorn R.R.A. (2006). Balloon crash damage and injuries: an analysis of 86 accidents, 2000-2004. *Aviation, Space, and Environmental Medicine 77*(5), 556-8.

Kobayashi A. (1996). Trace element and hormonal response during flight aptitude test. *Aviation, Space, and Environmental Medicine 67*(4), 333-337.

Leino T.K., Leppaluoto J., Ruokonen A., & Kuronen P. (1999). Neuroendocrine responses to psychological workload of instrument flying in student pilots. *Aviation, Space, and Environmental Medicine 70*(6), 565-570.

Li G., Baker S.P., Grabowski J.G., & Rebok G.W. (2001). Factors associated with pilot error in aviation crashes. *Aviation, Space, and Environmental Medicine 71*(1), 52-58.

Miller R.G. (1970). The stress of aircraft carrier landings. I. Corticosteroid responses in naval aviators. *Psychosomatic Medicine 32*(6), 581-588.

Munck A., Guyre P.M., & Holbrook N.M. (1984). Physiological functions of glucocorticoids. *Endocrine Review 5,* 25-44.

Sive W.J., & Hattingh J. (1991). The measurement of psychophysiological reactions of pilots to a stressor in a flight simulator. *Aviation, Space, and Environmental Medicine 62*(9), 831-836.

Ursin H., & Olff H. (1993). Psychobiology of coping and defence strategies. *Neuropsychobiology 28*(1-2), 66-71.

Educational Learning Theories: Informing the Fundamentals of Instruction

Amy L. Hoover
Department of Aviation
Central Washington University
400 E University Way
Ellensburg, WA 98926-7515
hoovera@cwu.edu

Abstract

The Federal Aviation Administration bases its fundamentals of instruction (FOI) primarily on principals of cognitive theory and behaviorism. However, other flight training curricula are centered on concepts and practices from social learning theory such as constructivism, mastery learning, mentoring, reciprocal teaching, and the integration of cognitive, affective, and psychomotor domains of learning and are more aligned with current practices in education. Thus some Flight Instructor applicants may pass the FOI knowledge exam without gaining a complete understanding of important underlying educational learning theories applicable to flight training. Examples from the educational literature are used to describe some of those social learning theories and relate them to design and delivery of flight training curricula to enhance the transition from theory to practice.

Educational Learning Theories: Informing
the Fundamentals of Instruction

Flight training is a kinesthetic, visual, cognitive, and often emotional learning experience; students bring expectations, doubts, and even fears with them into the cockpit. To facilitate learning flight instructors must be able to understand and relate relevant teaching and learning theories to training practices so as to build on the student's positive expectations and allay negative emotional influences. To that end, the Federal Aviation Administration (FAA) requires applicants for the Certified Flight Instructor (CFI) rating to pass a knowledge exam over the fundamentals of instruction (FOI). Although the intent is that flight instructors will apply appropriate learning theory to the design and delivery of pilot training courses, conversations among instructors and pilot examiners reveal a concern that CFIs are often ill prepared to do so. The FOI as presented by the FAA Aviation Instructor's Handbook are based heavily on concepts of cognitive theory and behav-

iorism (FAA, 1999). Other flight instructor training curricula have a broader basis in a variety of cognitive and social learning theories and practice, including constructivism, reciprocal teaching, mastery learning, mentoring, and the integration of cognitive, affective, and psychomotor domains of learning (Gleim, 2007; Jeppesen, 2002) and are more aligned with current theory and practice in education (Schunk, 2003),

This discussion describes several major learning theories not addressed in the FAA flight instructor training materials with examples from the educational literature as they relate to design and delivery of a variety of flight training curricula. The intent is to add a depth and richness to the overall understanding of learning theories and facilitate a transition from theory to application and practice in flight instruction.

Learning Theory Foundations for Fundamentals of Instruction

Social learning theories share the concept that humans learn through interaction with others and with their environment (Bandura, 1977; 1986, 1993; Schunk, 2003; Vygotsky, 1978, 1979). Thus, social learning is affected by the culture in which the individual is enmeshed and cognitive development results from shared experiences and interactions with individuals or groups that include both instructors and more competent peers. Jeppesen (2002) advocates a learning environment in which the instructor builds a trusting relationship with the student and utilizes concepts of social learning to enhance the student's progress throughout training. This emphasis on building trust is critical, because the student understands they are literally placing their life in the hands of the instructor during flight training; if the student does not trust the instructor learning can be compromised. When instructors work to build trust and guide students through their learning in a secure environment they are practicing the concepts of mentoring as described by Daloz (1986). The flight training method of "instructor does, instructor tells," "instructor does, student tells," and "student does, student tells" (FAA, 1999; Jeppesen, 2002) utilizes the concepts of reciprocal teaching, which is an applied method based on Vygotsky's (1978) theory of social cultural learning. It is an interactive process in which the teacher poses questions or models a skill and as the process continues, the student takes turns being the teacher.

Another important applied learning theory is the concept of constructivism. This theory posits that students construct their own knowledge though an active learning process based on existing beliefs and experiences (Schunk, 2003) Flight instructors are taught:

> Constructivism is based upon the idea that learners construct knowledge through the process of discovery as they experience events and actively seek to understand their environment. To employ Constructivism, your role shifts from the transmitter of information to the creator of experiences. (Jeppesen, 2002, p. 1-11)

When the flight student is able to construct their own experience they are better able to facilitate mastery learning. Mastery learning is the principle method of instruction incorporated into the building block method of integrated flight instruction. Described as a major component of social cognitive learning (Bandura, 1977;

1986), mastery learning is a process that incorporates goal setting in small incre- ments combined with continuous feedback, correction, and enrichment to assure that the learner masters concepts and skills before progressing to the next level (Bloom, 1974; Burton, Brown, & Fischer, 1984). Mastery learning is especially effective when learning extremely complex skills, as described by Burton et. al. (1984):

The student is exposed to a sequence of environments (microworlds) in which his tasks become increasingly complex. The purpose of an individual microworld is to provide the student with a task that he can perform successfully using a sim- plified version of the final skill that is that goal. This allows the student to focus on and master one aspect of the skill in a context that requires related subskills. As a result, the student learns when to use the skill as well as how to use it. Thus the purpose of the sequence is to evolve the simplified skills toward the goal skill. (p. 139)

A technique closely related to mastery learning called instructional scaffolding (Bruning, Schraw, & Ronning, 1995) involves a process by which the teacher controls the number of tasks to be learned and, based on the student's progress, introduces the next set of tasks or skills until the student is able to master those skills and move on to the next level.

The flight instructor literature stresses that all aviation-related training must go beyond rote levels of learning and that students must achieve higher levels of experience, application, and insight in each of three learning domains – cognitive, affective, and psychomotor (FAA, 1999; Gleim, 2007; Jeppesen, 2002). The cog- nitive domain involves knowledge and thought processes, and highest levels of learning include application, analysis, synthesis, and evaluation of those thought processes. Students exhibit higher levels of learning in the psychomotor domain through positive response, adaptation, and combining simple to more complex acts to form new movement patterns. The affective domain includes the student's attitudes, beliefs, and values; in order to achieve higher levels of learning in the affective domain students must be able to accept and create value in their life for the learning experience. Interestingly, the FAA instructor manual notes that "The affective domain may be the least understood, and in many ways, the most impor- tant of the learning domains" (FAA, 1999, p. 1-11). Thus, effective design and delivery of any flight training course must attend to how students are involved emotionally in the learning process. Such a course must also integrate educa- tional objectives that stimulate appropriate mental activity, states of mind, and demonstration of skills at the highest levels possible in each of the three learning domains. The relationship of underlying themes of social learning theory and strategies for their application to teaching and learning are discussed next.

Learning Theories Applied to Flight Instruction

Based on more than twelve thousand hours as a flight, classroom, and simu- lator instructor this author has anecdotal evidence that modeling and facilitation of mastery learning is a highly effective way to ensure positive transfer of learning.

The techniques of building on what the flight student already knows and continuously challenging them to master skills incrementally to achieve higher levels of performance has proven effective. Flight training curricula emphasize that students need time to practice maneuvers and skills to form their own understanding of relationships between theory and application (FAA, 1999; Gleim, 2007; Jeppesen, 2002). Focusing on the flight student's strengths to build their confidence and self-motivation toward mastering concepts and skills allows them to be able to make mistakes (as long as it is safe) to help them better analyze and synthesize their understanding of those mistakes and create their own solutions. This combination of techniques is founded in constructivism, mastery learning, and social learning theories.

There are many teaching applications based on social cognitive theory, which focuses on the relationship between learning and motivation of the individual within a social context. Bandura (1977; 1986; 1993) defined social cognitive learning as based on two principle concepts regarding the learner: 1) self-directed learning, which is the ability of a learner to set and achieve realistic and attainable goals for themselves; and 2) self-efficacy, or the learner's positive belief in their own ability to learn and master a concept or task and to achieve the goals they have set. Both concepts rely on a high degree of motivation, which Bandura describes with reference to two distinct domains: external motivations (e.g., rewards, praise, reinforcement, recognition) and internal motivations (e.g., past experiences, personality, desires, goals, curiosity, choices, and persistence). Internal motivations are most significant, while external motivations, although abundant, are generally less important to the learner (Bandura, 1986; 1993). In social cognitive theory, the internal domain dominates the learner's motivation. The flight student's concept of success is based on her own sense of personal achievement, belief in the significance of her contributions, and belief in her ability to achieve her goals. Those beliefs comprise the student's self-efficacy (Bandura, 1993). Figure 1 illustrates that relationship.
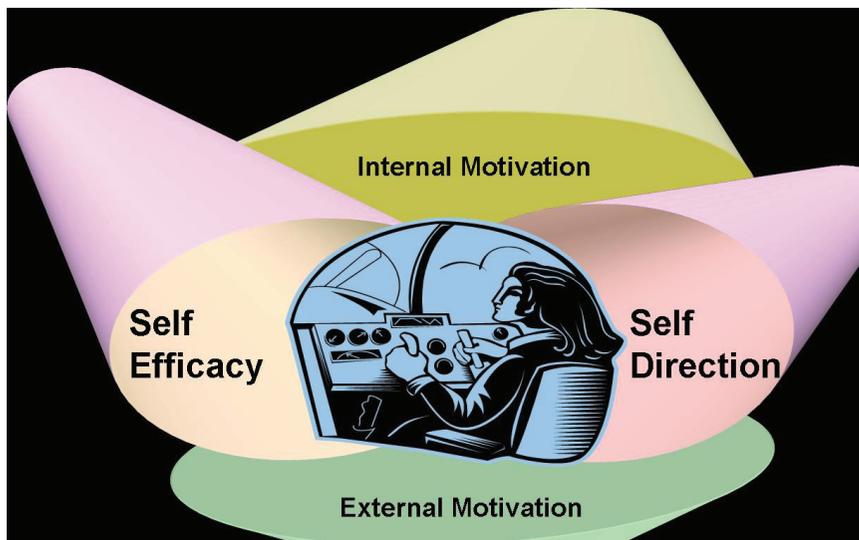

Figure 1. The relationship of internal and external motivations to the flight student as a social cognitive learner.

Based on social cognitive learning theory, any teaching strategy that enhances the individual's self-efficacy serves to increase her success in achieving her goals and learning objectives by giving her control over her own learning. The most significant strategies are:

- Mastery learning (Bloom, 1974; Bandura, 1993) and reciprocal teaching (Vygotsky, 1978) as previously discussed.
- Modeling, in which the flight student observes and interprets a behavior, then adopts that behavior if it has functional value or results in outcomes they value (Bandura, 1977; 1993; Decker & Nathan, 1985).

Social persuasion, in which the teacher or flight instructor serves as the model for mastery learning by facilitating self-efficacy and self-directed learning. The learner is more likely to adopt a modeled behavior if the mentor is similar to the learner and has admired status (Bandura, 1993). Jeppesen (2002) stresses this in its approach to flight instruction through emphasis on developing a common core of experience between flight instructor and student. To develop that relationship the instructor must place the student's learning as the primary lesson objective and maintain a professional, accommodating style while respecting and accepting the student.

Decker and Nathan (1985) based their behavior modeling concepts on the theoretical work of Bandura (1977) and restated his theory:

> In more informal terms, in order for people to learn from behavior modeling training, they must *observe* what the model is doing, *remember* what the model did, *do* what the model has done, and later when the appropriate time comes, *want to use* what they have learned. (p. 4)

Decker and Nathan (1985) incorporated those four concepts based on Bandura's theory into five strategies for behavior modeling training, which include 1) modeling, or the presentation or display of a behavior; 2) retention, which includes the learner's mentally practicing the behavior or coding it by writing it down or verbally describing it; 3) rehearsal, which includes the learner practicing the modeled behavior; 4) feedback, which is provided by the instructor or trainer and serves both as a constructive tool to improve performance and as a social reinforcement for the acceptance of the new behavior; and 5) transfer of training (which the FAA calls transfer of learning) in which the learner applies the newly acquired behavior in context (Decker & Nathan, 1985). Flight instructor training outlines each of these steps as important components of an effective teaching process (FAA, 1999; Jeppesen, 2002).

Burton, et al. (1984) used learning to ski as an example of modeling and mastery learning. The beginning skier models the behavior of the instructor, who helps him set small, realistic goals as intermediate levels of expertise that he can master incrementally. Because the learner sees the instructor excel at skills he has great desire to master, and the instructor gives him the tools and support necessary to progress and master each step, he feels confident to progress to the next more difficult task (self-efficacy). The environment also plays a key role in

ski instruction: for example, progressing from more gentle to steeper slopes incrementally helps the skier build confidence each time he masters the new slope. Similarly, introduction of the graduated length method (initial use of shorter skis, then graduating to longer and longer skis as skills progress) takes advanced ski instruction to a completely new level using this mastery learning technique.

Learning to fly is much like learning to ski because it is a complex skill in which the starting and final states are far apart. Additionally, the consequences of not mastering flying skills at many levels can be deadly from beginning to advanced skills such as instrument flying or aerobatics. Learners may become doubtful or even fearful if their belief in their ability to master skills is compromised. For this reason, the flight instructor must become a trusted mentor who builds and supports the student's sense of self-efficacy and facilitates mastery learning; the instructor must ascertain the student's internal motivations as they relate to the flight training objectives and build upon those motivations through positive reinforcement.

It is possible for a student to perform a basic rudimentary skill without fully understanding the relationships between the various flight parameters. For example, the student might perform a change of airspeed in level flight though memorization of proper control inputs or by experimentation with control inputs until the correct combination is achieved. However, in order to correlate those control inputs to performance in more complex flight regimes, such as change of airspeed in climbs, descents, turns during configuration changes, or any combination of those, the student must understand the underlying aerodynamic principles that govern the relationship between pitch, power, and airspeed. To assure the student is able to make those correlations the flight instructor can apply the basic practices of mastery learning. Because mastery learning is achieved by separating complex skills into simpler subskills to be mastered sequentially and progressively, the instructor should model a task (subskill) that the student can perform successfully and help the student set the goal of mastering each subskill in turn. For example, the instructor will help the beginning student to master the separate subskills of straight and level flight, climbs, descents, and level turns. Once the student gains confidence and proficiency in each of those maneuvers, those skills can then be combined to achieve mastery of climbing and descending turns. The process reinforces the student's belief in his ability to perform tasks as he continues to master each subskill. When the student has mastered a number of subskills, those skills can be fitted together as components of the more complex skills required in advanced maneuvers such as takeoffs and landings. By this method the student pilot will gain confidence in his ability to perform skills and insight into how to combine those skills in various flight operations. If the student attempts to master the more complex skills initially without first learning the requisite subskills he may become frustrated and learning can be compromised.

Some flight instructors advocate continued repetitive practice of a concept, skill, or maneuver even after the student has mastered the ability to perform to completion standards. Continuing to rehearse information after the initial learning is accomplished is called overlearning (Krueger, 1929). Overlearning is a technique that has been widely used, especially when learning highly intricate or technical skills such as playing a musical instrument or learning a language, and has

been tested often in the literature (Driskell, Willis, & Cooper, 1992). In a meta-analysis of 15 experimental studies that investigated the effectiveness of over-learning Driskell et. al. (1992) found that overlearning greatly increased performance, but the increase usually disappeared in a short period of time, typically between one and three weeks.

More recent research questions the validity of overlearning as an effective strategy (Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005). In two separate experimental studies they discovered that the boost in learning certain memory recall tasks mostly disappeared within nine weeks. Additionally, overlearning was highly inefficient; a quadrupling of study time produced far less than a doubling of recall rate when subjects were tested one week later (Rohrer et. al., in press). Overlearning strategies are typically used in flight training for rote memorization tasks and may be most effective at those lowest levels of learning, but not for mastery learning or scaffolded instruction necessary to teach more complex skills.

The teaching and learning strategies related here have a timely relevance to the FAA/Industry Training Standards (FITS) program introduced in 2002 to facilitate training in technically advanced aircraft. According to the FAA, FITS was designed to create "scenario-based, learner-focused training materials that encourage practical application of knowledge and skills," (FAA, 2002). Design of FITS training is similar to a type of training widely used in the airline industry known as line oriented flight training (LOFT) modeled after the original training implemented by Northwest Airlines in the late 1970's (Kern, 1998). In LOFT pilots fly a complete flight profile, including departure, enroute, arrival, and approach segments, in which specific training challenges have been predetermined and are introduced during the flight (e.g., systems and equipment malfunctions, traffic delays, weather hazards, diversions). For FITS, these types of training flights can be designed for simulated sessions or real-time training in aircraft. Instructors can model appropriate responses and behaviors as the student progresses through the decision making process in dealing with the LOFT challenges. Each challenge can be modified or repeated until the student attains mastery of appropriate sub-skills and then move to the next challenge. Through use of modeling, mastery learning, and scaffolded instruction students learning through scenario-based training should be able maximize their ability to learn at the highest levels in each of the learning domains.

Summary

Effective flight training must be solidly grounded in educational learning theory. It is important for flight instructors to understand that, although the FOI are based primarily on cognitive theory and behaviorism, there exists a large and well-documented body of research and literature in other, possibly more appropriate, learning theory. Flight instructor applicants need to gain a thorough working knowledge of concepts and teaching strategies based on relevant underlying learning theories from the educational literature. In order to do that, instructors must accept the basic tenets of FOI and value them as major components in the

successful design and delivery of flight training curricula. Additionally, they must be able to apply strategic training methods, such as those described in this discussion, to a variety of flight training arenas, including training for initial flight instructor applications. Through strategies such as modeling, mentoring, mastery learning, and social persuasion, flight instructors can facilitate the highest objective levels of learning in all learning domains.

References

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ. Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action.* Engelwood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117-148.

Bloom, B. S. (1974). An introduction to mastery learning theory. In J.H. Block (Ed.), *Schools, society, and mastery learning* (pp. 3-14). New York: Holt, Rinehart, & Winston.

Bruning, R. H., Schraw, G. J. & Ronning, R. R. (1995). *Cognitive psychology and instruction* (2nd ed.). Englewood Cliffs, NJ: Merrill/Prentice-Hall.

Burton R. R., Brown J. S., & Fischer, G. (1984). Skiing as a model of instruction. In B. Rogoff & J. Lave (Ed.), *Everyday cognition: Its development in social context* (pp. 139-150). Cambridge, Harvard University Press.

Daloz, L. (1986). *Effective teaching and mentoring: Realizing the transformational power of adult learning experiences.* San Francisco: Jossey-Bass.

Decker, P. J., & Nathan, B. R. (1985). *Behavior modeling training principles and applications.* New York, Praeger Publishers.

Driskell, J. E., Willis, R. P., & Cooper, E. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*(5)*,* 615-622.

Federal Aviation Administration (1999). *Aviation instructor's handbook* (FAA-H-8083-9). Washington D.C.: U.S. Government Printing Office.

Federal Aviation Administration, 2002. *FAA/Industry Training Standards (FITS). Questions and Answers.* Retrieved 11/5/08 from www.faa.gov/education_research/training/fits

Gleim, I. N., (2007). Flight Instructor Flight Maneuvers and Practical Test Prep. Gleim Publications, Inc. Gainsville, FL. 571 pp.

Jeppesen (2002). *Flight instructor manual*. Englewood, CO: Author.

Kern, T. (1998). *Flight discipline*. New York: McGraw-Hill.

Krueger, W. C. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12,* 71-78.

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005) The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19*(3), 361-374.

Schunk, D. H. (2003). *Learning theories: An educational perspective* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Vygotsky, L. (1979). Consciousness as a problem in the psychology of behavior. *Soviet Psychology*, *17*(4), 3-35.

Erratum

S. M. Casner, (2008) General Aviation Pilots' Attitudes toward Advanced Cockpit Systems, *International Journal of Applied Aviation Studies 8*(1), 88-113.

In Figure 2, on page 92, the bar chart displayed was incorrect. The correct figure is shown below. This figure has been corrected in the electronic version of the *International Journal of Applied Aviation Studies,* http://www.faa.gov/about/office_org/headquarters_offices/arc/programs/academy/journal/ and will be corrected in all future republications.
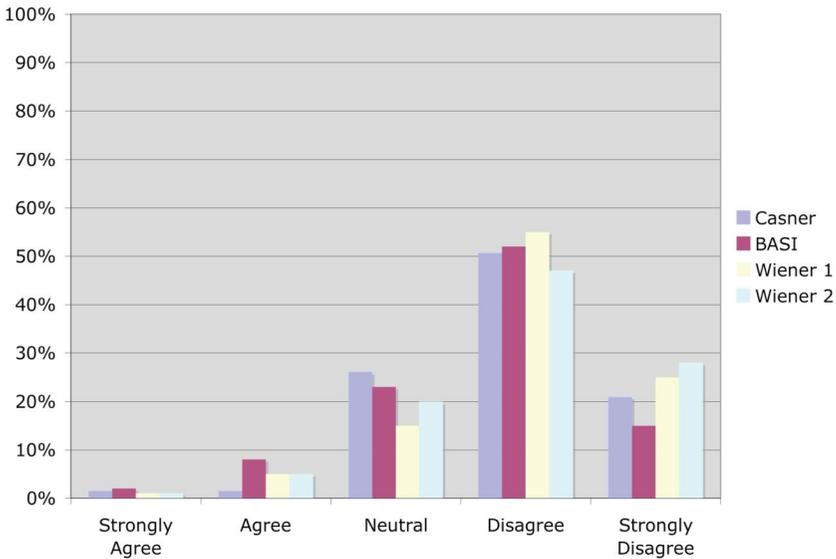

Figure 2. "They've gone too far with advanced cockpit systems."