

2006

International Journal of Applied Aviation Studies

Volume 6, Number 2

A publication of the FAA Academy



**Federal Aviation
Administration**

International Journal of Applied Aviation Studies

A Publication of the FAA Academy Oklahoma City, Oklahoma

Volume 6, Number 2, 2006
Government Printing Office

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.

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

International Journal of Applied Aviation Studies
Kay Chisholm
AMA-530C
PO Box 25082
Oklahoma City, OK 73125



International Journal of Applied Aviation Studies
Volume 6, Number 2
ISSN Number: 1546-3214
Copyright © 2006, FAA Academy
1st Printing December 2006

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
Deputy Superintendent

Gary Condley,
Robert Igo

PUBLICATION COMMITTEE

Manager, Regulatory Standards Division
Manager, Technical Operations Training Division
Manager, Air Traffic Division
Manager, Airports & International Training Division

David Long (Acting)
Jessie McMullen
Patricia Garrambone
Doug Andresen (Acting)

EDITORIAL STAFF

Managing Editor
Publication Editor
Production Editor
Associate Editor
Research Consultant

Kay Chisholm
Beverly Laughead
Laurie Wilson
Charlene Gahagan
Todd P. Hubbard

CONSULTING EDITORS

International

Ruwantissa I.R. Abeyratne	International Civil Aviation Organization, Montreal, Canada
Robert Bor	Royal Free Hospital, London, UK
Finian Connolly	Executive & Professional Training Institute, Drogheda, Co Louth, Ireland
Hugh David	Eurocontrol, East Sussex, UK
Sidney Dekker	Linköping Institute of Technology, Linköping, Sweden
Alois Farthofer	Human Factors Research, St. Georgen, Austria
Gerard J. Fogarty	University of Southern Queensland, Toowoomba, Australia
Jarle Gimmestad	Braathens Airlines, Fornebu, Norway
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
K. Wolfgang Kallus	Karl Franzens University, Graz, Austria
Michael John Kay	RMIT University, Melbourne, Victoria, Australia
Elizabeth Mathews	International Civil Aviation Organization, Montreal, Canada
Jens Rolfsen	Det Norske Veritas, Høvik, Norway
Steve Thatcher	University of South Australia, Mawson Lakes, Australia
Claudia van de Wal	VALK Foundation, Netherlands
Lucas van Gerwen	VALK Foundation, Netherlands
Stephen Walsh	Interaction Training Associates, Netherlands
Rogier Woltjer	Linköping Institute of Technology, Linköping, Sweden
Beat Zimmermann	Air Navigation Institute, Thun, Switzerland

U.S.A.

Dee H. Andrews	Air Force Research Laboratory, Mesa, AZ
R. Kurt Barnhart	Indiana State University, Terre Haute, IN
Ellen Bass	University of Virginia, Charlottesville, VA
Jeffrey 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
Aleta Best	Department of Transportation, Washington, DC
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
John Cain	Florida Technology Institute, School of Aeronautics, Melbourne, FL
Raymond E. Cain, Jr.	Florida Memorial University, Miami Gardens, FL
Michele Cappelle	FAA National Headquarters, Washington, DC
David Carmichael	Oklahoma City University, Oklahoma City, OK
Stephen M. Casner	NASA Ames Research Center, Moffett Field, CA
Robert H. Chisholm	Support Systems Associates, Melbourne, FL
Janet Conway	Southeastern Oklahoma State University, Durant, OK
Danet Cope	Washington Consulting Group, Prescott, AZ
Carmen Delgado-Morales	FAA Academy, Oklahoma City, Oklahoma
Brian G. Dillman	Purdue University, West Lafayette, IN
Key Dismukes	NASA Ames Research Center, Moffett Field, CA
Graham Elliott	FAA Academy, Oklahoma City, OK
Richard Fanjoy	Purdue University, West Lafayette, IN
Douglas R. Farrow	Federal Aviation Administration, Washington, DC
Kenneth H. Funk II	Oregon State University, Corvallis, OR
Trent Gabert	University of Oklahoma, Norman, OK
Patricia Garrambone	FAA Academy, Oklahoma City, Oklahoma
A. F. Grandt	Purdue University, West Lafayette, IN
Mavis Green	Harvard University, Cambridge, MA
Carla A. Hackworth	Civil Aerospace Medical Institute, Oklahoma City, OK
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
Judy Holcomb	FAA Academy, Oklahoma City, OK
Amy Hoover	Central Washington University, Ellensburg, Washington
Todd P. Hubbard	Oklahoma State University, Stillwater, OK
Michael Hynes	Hynes Aviation Services, Frederick, OK
Sig Illing	FAA Academy, Oklahoma City, OK
Bill Johnson	FAA, Aviation Safety, College Park, GA
Merrill R. Karp	Arizona State University, Mesa, AZ
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, Oklahoma
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

Richard L. Mangrum	Kent State University, Kent, OH
Carol Manning	Civil Aerospace Medical Institute, Oklahoma City, OK
Harry Marks	University of Oklahoma, Norman, 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
H. C. McClure	MAC Air Consulting, Queen Creek, AZ
Marjo Mitsutomi	University of Redlands, Redlands, CA
Richard H. Mogford	NASA Ames Research Center, Moffett Field, CA
Henry Mogilka	FAA Academy, Oklahoma City, OK
John C. Morey	Dynamics Research Corporation, Andover, MA
Kathleen Mosier	San Francisco State University, San Francisco, CA
Vahid Motevalli	Aviation Institute, Ashburn, VA
Gary J. Northam	Parks College, St. Louis University, St. Louis, MO
Dale B. Oderman	Purdue University, West Lafayette, IN
Manoj Patankar	Parks College, St. Louis University, St. Louis, MO
Mike Pearce	American Airlines
Donald A. Petrin	Purdue University, West Lafayette, IN
Julia Pounds	Civil Aerospace Medical Institute, Oklahoma City, OK
Roni Prinzo	Civil Aerospace Medical Institute, Oklahoma City, OK
Stephen M. Quilty	Bowling Green State University, Bowling Green, OH
William B. Rankin II	Florida Memorial College, Miami, FL
Esa M. Rantanen	University of Illinois, Urbana-Champaign, IL
Gail Reed	Pan Am International Flight Academy, Miami, FL
Sherry Reese	FAA Academy, Oklahoma City, OK
Robert A. Richards	Stottler Henke Associates, Inc. (SHAI), San Mateo, CA
Robert Ripley	Auburn University, Auburn, AL
Charles L. Robertson	University of North Dakota, Grand Forks, ND
Nadine Sarter	The Ohio State University, Columbus, OH
Robert Schlegel	University of Oklahoma, Norman, OK
Gregory L. Schwab	Indiana State University, Terre Haute, IN
John W. Sheremeta, Jr.	American Airlines, Lake Ronkonkoma, NY
Mark A. Sherman	Ohio University, Athens, OH
Edmond Soliday	Flight Safety Foundation, Alexandria, VA
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
Craig Gerald Stroup	American West Airlines, Phoenix, AZ
Donald Anders Talleur	University of Illinois, Urbana-Champaign, IL
Scott Tarry	Aviation Institute, University of Nebraska, Omaha, NE
Bill Traylor	Department of Transportation, Fort Worth, TX
Ron Ward	FAA Academy, Oklahoma City, OK
Thomas Weitzel	Embry-Riddle Aeronautical University, Daytona Beach, FL
Michael Wiggins	Embry-Riddle Aeronautical University, Daytona Beach, FL
Kevin Williams	Civil Aerospace Medical Institute, Oklahoma City, OK
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.

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

EDITOR'S NOTES

Formal Papers

The International Journal of Applied Aviation Studies' lead article is by Raymond King, Carol Manning, and Gena Drechsler. The lead article position is awarded to the highest scoring submission from the peer review. *Operational Use of the Air Traffic Selection and Training Battery* presents a study to determine how the reweighing effort, from a previous project, fares with actual applicants in the goal of reducing/eliminating group differences that could result in potential unfair discrimination. In addition, the study examines how this reweighing functioned operationally to eliminate group differences.

To better understand the current state of the ASAP programs, *A Review of the Current State of Aviation Safety Action Programs in Maintenance Organizations* by Patankar and Ma, presents the results of a survey of 20 maintenance organizations. Based on the results of their study, the authors argue that the effectiveness of ASAP programs in maintenance should be measured by the percentage of actual changes at each of the three levels of impact, rather than the number of total or sole-source reports.

In *Flight Crew Callouts and Aircraft Automation Modes*, Goteman and Dekker aimed to augment previous research on flight crew callouts conducted in simulator settings with natural observations of real cockpit activity. Their results provide answers to whether Flight Mode Annunciator (FMA) callouts are used as a tool to detect and remember automation mode changes, or as a vehicle for coordinating between pilots.

Schulman explores the link between an airline's profitability and its safety record in *Financial Stability and Airline Safety: Relationships, Causes, and Consequences*. The purpose of this study was to determine the relationship between profitability and safety. Schulman investigates the extent that investments in safety projects and the level of maintenance outsourcing impact safety.

In *A Case-based Review of Critical Incidents in General Aviation for Improved Safety*, Saleem and Kleiner report on critical incidents in which pilot error occurred during field observations of landing approaches to a mid-sized, controlled airport. Each of the incidents involved communications errors, where information exchanged from ATC to the pilot was not initially processed by the pilot or the information exchanged was incorrect. The authors use these incidents to demonstrate how enhancements may help minimize the occurrence of similar errors.

Enhancing Life in the Hyper-Surveillance Mini-World of a Space Station: The Role of Situation Awareness, Communication, and Reality TV in the Life of Astronauts is the third article of a series entitled *Astronauts as Audiences* by Rankin and Cokley. This article investigates the roles that situation awareness (SA), communications, and reality TV might have on the lives of astronauts in remote space

communities and surmises what the collective application of these roles might be as a means of enhancing the lives of astronauts in remote space communities.

Macchiarella, Arban, and Doherty conducted an 18-month study that applied an experimental flight-training curriculum to certify Private Pilots under Federal Aviation Regulation (FAR) Part 142. In *Transfer of Training from Flight Training Devices to Flight for Ab-Initio Pilots*, the authors present their study and discuss the results.

The Evaluation of the Decision Making Processes Employed by Cadet Pilots Following a Short Aeronautical Decision-Making Training Program. In this study, Li and Harris developed a short, ADM training course constructed around two mnemonic methods, SHOR (Stimuli, Hypotheses, Options, and Response) and DESIDE (Detect, Estimate, Set safety objectives, Identify, Do, Evaluate). After the training, the procedural knowledge underpinning their Situation Assessment and Risk Management ability were evaluated using knowledge tests based upon several demanding tactical flight situations. The authors discuss the results and their implications.

Book Reviews

Developing Strategies for the Modern International Airport, East Asia and Beyond, by Allan Williams, focuses on the aviation industry's role in the development of the international economy. John C. Di Renzo Jr. gives a comprehensive review of this book that also covers past and current aviation industry issues

Todd P. Hubbard gives the reader a vivid review of *What We Should Know About Human Error: A Review of Ten Questions about Human Error* by Sidney Dekker. Dekker looks at the problems in the psychological study of aviation--the causes and solutions.

KC

International Journal of Applied Aviation Studies

Volume 6, Number 2	2006
Review Process	197
Policy and Disclaimer	198
Production Staff	199
Philosophy Statement	202
Editor's Notes	203
Formal Papers	
Operational Use of the Air Traffic Selection and Training Battery <i>Raymond E. King, Carol A. Manning, and Gena K. Drechsler</i>	207
A Review of the Current State of Aviation Safety Action Programs in Maintenance Organizations <i>Manoj S. Patankar and Jiao Ma</i>	219
Flight Crew Callouts and Aircraft Automation Modes An Observational Study of Task Shedding <i>Örjan Goteman and Sidney Dekker</i>	235
Financial Stability and Airline Safety: Relationships, Causes, and Consequences <i>Adam B. Schulman</i>	249
A Case-based Review of Critical Incidents in General Aviation for Improved Safety <i>Jason J. Saleem and Brian M. Kleiner</i>	271
Enhancing Life in the Hyper-Surveillance Mini-World of a Space Station: The Role of Situation Awareness, Communication, and Reality TV in the Life of Astronauts <i>William B. Rankin and John Cokley</i>	283
Transfer of Training from Flight Training Devices to Flight for Ab-Initio Pilots <i>Nickolas D. Macchiarella, Pamela K. Arban and Shawn M. Doherty</i>	299
The Evaluation of the Decision Making Processes Employed by Cadet Pilots Following a Short Aeronautical Decision-Making Training Program <i>Wen-Chin Li and Don Harris</i>	315
Book Reviews	
Developing Strategies for the Modern International Airport East Asia and Beyond by Allan Williams <i>Reviewed by John C. Di Renzo Jr.</i>	335
What We Should Know About Human Error: A Review of Ten Questions about Human Error by Sidney Dekker <i>Reviewed by Todd P. Hubbard</i>	339

Formal Papers

Operational Use of the Air Traffic Selection and Training Battery

Raymond E. King

Carol A. Manning

and

Gena K. Drechsler

*Civil Aerospace Medical Institute
6500 South MacArthur
Oklahoma City, OK 73169, USA*

Abstract

The Federal Aviation Administration (FAA) is commencing a massive hiring of air traffic control specialists using a new selection procedure, the Air Traffic Selection and Training (AT-SAT) computerized test battery. Before AT-SAT could be used for hiring purposes, however, the issue of its potential for adverse impact (potential unfair discrimination) had to be addressed. A previous project (Wise, Tsacoumis, Waugh, Putka, & Hom, 2001) reweighed the subtests and adjusted the overall constant to mitigate potential group differences that could result in adverse impact, without unduly compromising validity. A subsequent study (Dattel & King, 2006) used research participants and found that this effort appeared to have achieved its goal of mitigating group differences that could result in adverse impact. The present study endeavors to: 1) describe how AT-SAT functions as an operational selection method with respect to the several applicant pools, and 2) determine how the reweighing effort fares with actual applicants in the goal of reducing/eliminating group differences that could result in adverse impact. Of the 854 applicants who have taken AT-SAT as part of a job application process (rather than as according to a research protocol), 219 applicants (25.64%) voluntarily disclosed their race; gender was known for 253 (29.63%). The results suggested that the reweighing effort is paying dividends as group differences that could result in adverse impact are not in evidence. While the initial numbers reported here are relatively small, the issue of group differences that could result in adverse impact will be continually monitored. Longitudinal validation, comparing AT-SAT results to training and on-the-job performance, is a research priority due to concerns about the overall passing rate of 93.33%, which is higher than the expected passing rate of 67%.

Operational Use of the Air Traffic Selection and Training Battery

A Plan for the Future: The FAA's 10-Year Strategy for the Air Traffic Control Workforce was submitted to the U.S. Congress in December 2004. This report has been updated and is now referred to as A Plan for the Future (FAA, 2006). The report provided the outline to mitigate anticipated controller retirements and contemplated strategies to achieve appropriate staffing levels, including hiring over 11,800 controllers over the next decade. The present paper focuses on the current status of the recently¹ implemented Air Traffic Selection and Training (AT-SAT) selection test battery, focusing specifically on the functioning of the reweighed AT-SAT and the potential for group differences that could result in adverse impact.² The data in this paper represent the first time operational (collected for selection purposes rather than research) AT-SAT data are being reported.

How did the staffing situation become so urgent? As detailed in the Controller Workforce Plan, a majority of the air traffic controller workforce went on strike on August 3, 1981, when President Ronald Reagan ordered the striking controllers to return to duty within 48 hours. When 10,438 (out of a workforce of approximately 15,000) striking controllers did not return to work by the deadline, President Reagan fired them. Facing a sudden shortage of controllers, the Federal Aviation Administration (FAA) hired 3,416 individuals in 1982 and another 1,720 in 1983. From 1982 through 1991, the FAA hired an average of 2,655 individuals per year. Hiring was much less robust in the early part of the first decade of the twenty-first century. The majority of entrants met the 18 to 30 years-of-age entry requirement.

The post-strike hiring wave created the potential for a large portion of the controller workforce to reach retirement age at roughly the same time, particularly due to the FAA policy requiring retirement from controlling air traffic by age 56. Based on current projections, about 70% of the agency's 10,500 controllers will become eligible to retire within ten years. Total losses are expected to reach nearly 10,300 (FAA, 2006). This amount of attrition means that the FAA must use effective recruitment, selection, and training procedures to ensure that its staffing needs are met. The AT-SAT battery, now the official Civil Service test used to select FAA air traffic control specialists (ATCSs) without previous operational air traffic control experience, will thus become an instrument of increasing importance.

The development and validation of AT-SAT plays a critical role in reducing costs associated with attrition from air traffic control training. Using a valid selection test also ensures that those who are hired have (or have the potential to develop) the necessary knowledge, skills, and abilities to be successful. The duties of ATCSs make them individually responsible for more lives than the practitioners of any other occupation in the United States (Biggs, 1979). The FAA developed the AT-SAT battery to replace a two-stage selection process in which ATCS applicants completed an Office of Personnel Management (OPM) test battery and a nine-week screening program at the FAA Academy in Oklahoma City, OK. This previous selection process proved to be expensive (Ramos, Heil, &

Manning, 2001a). AT-SAT was developed based on the results of the Separation and Control Hiring Assessment (SACHA; Nickles, Bobko, Blair, Sands, & Tartak, 1995) job task analysis, which drew heavily from previous work done by Ammerman, Becker, Jones, Tobey, and Phillips (1987). Additional development and validation efforts for AT-SAT were described by Ramos, Heil, and Manning (2001a & 2001b) in their two-volume report.

AT-SAT is a select-in procedure; select-out (medical) issues will not be addressed in this paper. Readers interested in select-out issues are referred to King, Retzlaff, Detwiler, Schroeder, and Broach (2003). AT-SAT is a computerized test battery comprised of eight subtests based on 22 individual scores that, when weighted (forming “part scores”) and combined, are totaled (with a constant added) for a single overall score. As delineated in Table 1, AT-SAT is comprised of the following subtests: Air Traffic Scenarios Test, ATST; Analogies, AY; Angles, AN; Applied Math, AM; Dials, DI; Experiences Questionnaire, EQ; Letter Factory, LF; and Scan, SC. AT-SAT is an aptitude test and not a test of air traffic control knowledge. The goal of AT-SAT is to predict the likelihood of success in air traffic control training and, more importantly, subsequently on the job. Seven of the eight subtests assess aspects of cognitive ability, while one, EQ, assesses issues in the personal history/personality realm. Four (ATST, AY, LF, SC) of the subtests are dynamic; they are interactive and can only be administered via computer. The remaining four are static, similar to pencil-and-paper tests, but are administered via computer.

Table 1
The Eight AT-SAT Subtests

Subtest	Description
Dials (DI)	Scan and interpret readings from a cluster of analog instruments
Applied Math (AM)	Solve basic math problems as applied to distance, rate, and time
Scan (SC)	Scan dynamic digital displays to detect targets that regularly change
Angles (AN)	Determine the angle of intersecting lines
Letter Factory (LF)	Participate in an interactive dynamic exercise that requires categorization skills, decision making, prioritization, working memory (incidental learning), and situation awareness
Air Traffic Scenarios (ATST)	Control traffic in interactive, dynamic low-fidelity simulations of air traffic situations requiring prioritization
Analogies (AY)	Solve verbal and nonverbal analogies that require working memory and the ability to conceptualize relationships
Experience Questionnaire (EQ)	Respond to Likert scale questionnaire about life experiences

During the development of AT-SAT, subtests were weighted to yield the maximum validity, according to their relationship to the job, as suggested by the job task analysis. An individual's performance on each subtest is multiplied by the weight developed for that subtest, and a constant is subtracted to ensure that each subtest has a floor of zero. The net result is a "part score." Part scores are then summed and combined with an additional, overall constant (composed of a summation of the inverse of the individual constants and a calibrating constant) to yield the overall AT-SAT score (which is truncated to have a maximum score of 100). Veteran's preference points (either five or ten points) are subsequently added. This overall score is the only one that enters into the hiring decision when AT-SAT is used in an operational fashion. Subtest weights are not disclosed in this paper as they could be used as part of a coaching strategy to artificially inflate AT-SAT scores in an effort to gain a competitive advantage.

Before operational use of AT-SAT was approved for hiring purposes, FAA employees that were members of minority groups raised concerns over potential adverse impact. Recall that adverse impact is a selection rate for any race, sex, or ethnic group that is less than $\frac{4}{5}$ (80%) of the rate for the group with the highest rate. To calculate this ratio, the pass rate of the group of interest serves as the numerator, while the pass rate of the group with the highest pass rate serves as the denominator. Eighty percent serves as the bright line; a quotient below 80% suggests a potential for adverse impact, while one equal to or greater than 80% argues against it. The concern about the potential for adverse impact against African Americans seemed well founded, as only three out of every 100 black applicants were predicted to achieve a score of at least 70 (the minimum passing score – termed a "qualifying score") on AT-SAT. Nevertheless, as the predictive validity was .69, a case could be made for "business necessity." Business necessity is a defense available when the employer has a criterion for selection that is facially neutral but excludes members of one sex, race, or national origin at a substantially higher rate than members of other groups (thus creating adverse impact). The employer must then prove that its selection requirement having the adverse impact is job-related, typically as demonstrated by a job analysis.

To appreciate the wider context, the reader needs to understand that the issue went beyond pass rates for minority applicants. By original design, 38% of fully certified incumbent FAA controllers would not pass AT-SAT under the original scoring scheme. The original passing score of 70 had been calibrated, using the overall constant, so that only 62% of incumbent fully certified controllers would achieve an AT-SAT score of 70 or more, in an effort to minimize FAA Academy failures and compensate for the need for ATCSs to perform potentially more difficult duties in the future. The goal was to at least preserve and strive to improve the level of functioning in the workforce (Waugh, 2001).

In response to concerns about group differences that could result in adverse impact, FAA officials requested that scientists review the options to mitigate these differences. Meanwhile, they emphasized maintaining the overall validity of the AT-SAT battery. Additionally, FAA management made the case that the cut score should be set at the point where most fully qualified, incumbent controllers would pass FAA's entry-level aptitude test. Consequently, the AT-SAT subtests were re-

weighted, and the overall constant was adjusted. The new weights were developed as a trade-off between their validity and their contribution to group differences that could lead to adverse impact. The content of the subtests themselves was not changed; rather, the subtests were weighted differently. The goal was to retain adequate validity while reducing potential adverse impact. Test validity (job-relatedness) is determined by the strength of the correlation between the overall AT-SAT score and en route controller job performance measures (which, in the AT-SAT validation study, consisted of a computerized situational judgment test assessing maximum performance, plus peer and supervisor ratings of typical performance). After reweighing, the correlation between AT-SAT and a job performance composite measure was reduced slightly, from .69 to .60 (Wise, Tsacoumis, Waugh, et al., 2001). Compared with most validation coefficients, this is still a strong correlation with job performance. As indicated above, the relationship with job performance is especially important in this context; any remaining adverse impact can be justified by business necessity, as previously defined above.

Using data from the original validation studies, Wise et al. (2001) found that, by reweighing the subtests and adjusting the overall constant (as described above), they eliminated group differences that could result in potential adverse impact for women and Hispanics and greatly reduced it for African Americans. Wise, et al. concluded their report with a cautionary statement about the uncertainty of how the reweighing might function with actual applicants, to include the impact on overall pass rates. A primary purpose of the present paper is to examine how this reweighing functioned operationally to eliminate group differences that could result in adverse impact.

To further address the potential problem of adverse impact, FAA officials decided to abandon a strict “top-down” approach to hiring and, instead, use a category ranking method. This approach is a form of “score banding” that can be justified on the basis of ignoring score differences that are due to an estimate of the applicant’s true ability. Score banding, although somewhat controversial among selection scientists, “will almost always produce less adverse impact than strict rank ordering” (Biddle, 2005, p. 103) as it ignores score differences likely to be statistically insignificant. Under this scheme, job fair applicants who achieve a qualifying minimum score are divided into two groups: those scoring 85 and above (termed “well qualified”) and those scoring from 70 to 84.9 (termed “qualified”). Those in the “well-qualified” group will be offered employment before anyone in the “qualified” group. AT-SAT data in this paper are therefore reported according to these categories.

Dattel and King (2006) applied the weights and additive constant developed to address potential adverse impact to the scores of 724 developmental ATCSs, hired by a previous method, who volunteered to take AT-SAT. An average increase of 4.86 points was found with the new scoring method; the notional passing rate, based on achieving an AT-SAT score of 70 or more, changed from 58.8% to 80%. American Indian/Alaskan Native, Hispanic, and black participants (categories are as defined on the Race and National Origin Form) showed the greatest average increase in overall scores: 6.97, 6.98, and 7.02 respectively, representing increases in pass rates of 22.2% (77.8% to 100%), 35.1% (51.9% to 87%), and 35.2% (37% to 72.2%), respectively.

While this analysis of data collected from research participants was encouraging, the real standard is to determine how a selection instrument functions with actual applicants. That opportunity has presented itself with the commencement of increased hiring and universal use of AT-SAT to hire candidates without previous air traffic control experience.

Sources of ATCS Applicants without Previous Controlling Experience

The Air Traffic - Collegiate Training Initiative (AT - CTI) program and job fairs are two authorities used to hire personnel without previous experience controlling air traffic. The AT - CTI program is a partnership between the FAA and 14 aviation colleges/universities to provide the academic preparation necessary for students interested in ATCS careers. Enrollment in, and even completion of, an AT - CTI program in no way obligates the FAA to hire the student, and the FAA typically does not test these students until near the completion of their studies. AT - CTI school officials decide whether to recommend a student to the FAA for AT-SAT testing, and hence, hiring consideration.

Not all job fair applicants are permitted to take the AT-SAT due to the time required (a full day) and the cost (about \$800) of administering it. Selection for AT-SAT testing was previously based on a random selection process but is now based on responses to a biographical questionnaire. Job fair applicants compete with each other in order to be hired at a specific facility. Not every job fair applicant who achieves a qualifying score will be offered employment. Consequently, it is possible that a job fair applicant who scores only in the "qualified" range will not be offered employment. Similarly, AT - CTI applicants who score in the "well qualified" range are referred for positions before applicants who score in the "qualified" range.

The third pool of applicants also did not have previous air traffic control experience, but were unique in that they took both the paper and pencil Office of Personnel Management (OPM) test and AT-SAT. Members of this group had passed the OPM test several years previously but were not admitted into the nine-week screening program due to reduced staffing requirements at the time. Since hiring was reduced after they passed the OPM test, but before they could complete the screening program, they had to take AT-SAT at a later time and could not be hired without a passing AT-SAT score (≥ 70). This group is also unique due to the amount of time that had elapsed before they were finally considered to be hired and their two tiers of testing.

Other Applicant Pools

There are other applicant pools for ATCS positions, including former military and Department of Defense civilian controllers. These groups will not be described in detail in this paper as they do not have to take AT-SAT and are considered for employment based on their previous operational experience controlling air traffic.

Method

Operational results (i.e., AT-SAT scores collected from applicants who competed on the basis of taking AT-SAT) were analyzed in terms of race, ethnic, and gender group membership, considered by hiring authority.

Participants

From June 2002 to June 2006, 854 applicants took AT-SAT as part of their job application process. Applicants either 1) responded to a job fair announcement (soliciting applicants for a specific position), 2) were nearing completion of one of the 14 AT - CTI programs, or 3) had previously passed the OPM test and had to achieve a passing score on AT-SAT before they were admitted into training at the FAA Academy. All applicants were requested to voluntarily complete the Race and National Origin (RNO) form (OPM Form 1468 until October 2005, FAA Form 3330-64 thereafter) to ascertain their racial/ethnic group membership. Self-identification of gender was new to FAA Form 3330-64; previously, that information was solicited through a variety of means. Of 854 job applicants, gender was known for 253 (29.63%). RNO data are discussed in the Results section, as they derive directly from the self-report forms.

Results

Of the 854 job applicants, 219 (25.64%) disclosed their race on the form in use at the time of their application. Total counts for race/national origin groups and gender are depicted in Table 2.

Table 2
RNO and Gender Results

RNO	Count
1.) American Indian or Alaska Native	1
2.) Asian (Asian or Pacific Islander previous to Oct 2005)	12
3.) Black or African American (Black, not of Hispanic Origin previous to Oct 05)	23
4.) Hispanic or Latino (Hispanic previous to Oct 05)	8
5.) Native Hawaiian or Other Pacific Islander (Was represented under #2 previous to Oct 05)	1
6.) White	174
Gender	Count
Male	212
Female	41

Table 3 displays AT-SAT pass rates (those who achieved a score ≥ 70) with respect to self-reported RNO group membership, segmented by hiring authority.

Table 3
RNO Group AT-SAT Differences, Considered by Hiring Source

Self-reported RNO Membership	Applicant Pool	Failed - scored below 70		Qualified - scored 70-85		Well Qualified – scored > 85		Total Passing – ? 70	
		Count	Row %	Count	Row %	Count	Row %	Count	Row %
1.) American Indian or Alaska Native	OPM/AT-SAT					1	100.0%	1	100.0%
2.) Asian	AT - CTI			2	66.7%	1	33.3%	3	100.0%
	Job Fair	1	14.3%	3	42.9%	3	42.9%	6	85.7%
	OPM/AT-SAT					2	100.0%	2	100.0%
3.) Black or African American	AT - CTI	2	14.3%	6	42.9%	6	42.9%	12	85.7%
	Job Fair	3	42.9%	4	57.1%			4	57.1%
	OPM/AT-SAT			2	100.0%			2	100.0%
4.) Hispanic or Latino	AT - CTI	1	20.0%	4	80.0%			4	80.0%
	Job Fair			1	33.3%	2	66.7%	3	100.0%
5.) Native Hawaiian or Other Pacific Islander	Job Fair					1	100.0%	1	100.0%
6.) White	AT - CTI	3	3.9%	20	26.0%	54	70.1%	74	96.1%
	Job Fair	2	3.2%	14	41.2%	20	58.8%	34	100.0%
	OPM/AT-SAT			20	31.7%	41	65.1%	61	96.8%
Total		12	5.5%	76	34.7%	131	59.8%	207	94.5%

Collapsing across applicant pools, the sole American Indian/Alaska Native applicant passed and the sole Native Hawaiian/Other Pacific Islander passed; 11 of 12 (91.67%) Asian applicants passed; 18 of 23 (78.26%) black applicants passed; 7 of 8 (87.5%) Hispanic applicants passed; and 169 of 174 (97.13%) white applicants passed. Ignoring the two smallest groups (due to their extremely small n), the ratio of the group with the lowest passing rate (black/African American) compared to the group with the highest passing rate is 80.57 (78.26/97.13), just above the 80% threshold, suggesting that AT-SAT did not, as yet, exhibit group differences that could result in adverse impact.

As the concept of adverse impact is also concerned with gender differences, the data were also examined by gender, again according to hiring authority. Table 4 displays AT-SAT pass rates (those who achieved a score ≥ 70) with respect to gender, segmented by hiring authority.

Table 4
Gender AT-SAT Differences, Considered by Hiring Source

Gender	Applicant Pool	Failed scored below 70		Qualified scored 70-85		Well Qualified scored > 85		Total Passing > or = 70	
		Count	Row %	Count	Row %	Count	Row %	Count	Row %
Male	AT - CTI	4	4.5%	30	34.1%	54	61.4%	84	95.5%
	Job Fair	2	4.8%	16	38.1%	24	57.1%	40	95.2%
	OPM/AT-SAT	3	3.7%	19	23.2%	60	73.2%	79	96.3%
Female	AT - CTI	3	13.6%	6	27.3%	13	59.1%	19	86.4%
	Job Fair			3	75.0%	1	25.0%	4	100.0%
	OPM/AT-SAT			7	46.7%	8	53.3%	15	100.0%
Total		12	4.7%	81	32.0%	160	63.3%	241	95.3%

Collapsing applicant pools, 203 of 212 men (95.75%) passed and 38 of 41 women (92.68%) passed. The ratio here is 96.79 (92.68/95.75), suggesting that AT-SAT did not exhibit a gender difference with a potential for adverse impact.

Table 5 delineates the results of the overall population of 854, by hiring source. The overall AT-SAT pass rate of 93.3% rate is higher than the 67% predicted in the Controller Workforce Plan (FAA, 2006).

Table 5
Overall AT-SAT Results, Considered by Hiring Source

Applicant Pool	Failed scored below 70		Qualified scored 70-85		Well Qualified scored > 85		Total Passing > or = 70	
	Count	Row %	Count	Row %	Count	Row %	Count	Row %
AT - CTI	34	6.9%	178	35.9%	284	57.3%	462	93.1%
Job Fair	20	7.7%	109	41.8%	132	50.6%	241	92.3%
OPM/AT-SAT	3	3.2%	26	26.8%	68	70.1%	94	96.9%
Total	57	6.7%	313	36.7%	484	56.7%	797	93.3%

Discussion and Conclusion

After years of validation efforts (Ramos, Heil, & Manning, 2001a & 2001b), AT-SAT is finally being used as an operational selection tool for ATCSs, coinciding with the increased hiring that has already begun. This study focused primarily on AT-SAT's performance in selecting applicants of various racial/ethnic backgrounds. The outcome found in this study is very different from the 3%

passing rate predicted under the original weighting scheme for black applicants, for example. There was also no evidence of any potential adverse impact with respect to gender. Therefore, while AT-SAT appears to be functioning as projected by Wise et al. (2001), in terms of reduced group differences that could result in adverse impact, the higher than expected pass rates may be of concern. The next step is to conduct longitudinal validation to determine how well AT-SAT predicts success in training and on-the-job performance.

The increased ATCS hiring will serve to populate the RNO and gender cells if applicants voluntarily provide the data. Group differences that could result in adverse impact will be continually monitored. The value of considering the data by hiring source is to suggest strategies to mitigate group differences, should they arise. Recall that AT-SAT scores are used differently depending on the hiring authority under which an applicant applies. Therefore, even though the present study is more realistic than previous efforts that used research data, a more thorough understanding could be afforded by an examination of applicants who are actually offered employment. In an effort to increase the voluntary participation rate, future applicants should be encouraged to complete their RNO forms so that a more accurate picture of the applicants can be achieved and group differences can be more readily detected. It will also be necessary to assess the potential for adverse impact of subsequent selection procedures, to include considering who is actually offered employment.

References

- Ammerman, H.L., Becker, E.S., Jones, G.W., Tobey, W.K., Phillips, M.D. (1987). FAA ATC operations concepts, *Volume I: ATC background and analysis methodology*. (DOT/FAA/AP-87-01). Washington, DC: Federal Aviation Administration.
- Biddle, D. (2005). *Adverse impact and test validity: A practitioner's guide to valid and defensible employment testing*. Burlington, VT: Gower.
- Biggs, D. (1979). *Pressure cooker*. New York: Norton.
- Dattel, A.R., King, R.E. (2006). *Reweighting AT-SAT to mitigate group score differences* (DOT/FAA/AM-06/16). Washington, DC: FAA Office of Aerospace Medicine.
- FAA (2006). *A Plan For The Future 2006-2015*. Retrieved September 30, 2006 from http://www.faa.gov/airports_airtraffic/air_traffic/controller_staffing/media/workforce_plan_060006.pdf
- King, R.E., Retzlaff, P.D., Detwiler, C.A., Schroeder, D.J., Broach, D. (2003). *Use of personality assessment measures in the selection of air traffic control specialists*. (DOT/FAA/AM-03/20). Washington, DC: FAA Office of Aerospace Medicine.
- Nickels, B.J., Bobko, P., Blair, M.D., Sands, W.A., Tartak, E.L. (1995). *Separation and control assessment (SACHA) final job analysis report*. Bethesda, MD: University Research.
- Ramos, R.A., Heil, M.C., Manning, C.A. (2001a). *Documentation of validity for the AT-SAT computerized test battery, Volume I*. (DOT/FAA/AM-01/05). Washington, DC: FAA Office of Aviation Medicine.
- Ramos, R.A., Heil, M.C., Manning, C.A. (2001b). *Documentation of validity for the AT-SAT computerized test battery, Volume II*. (DOT/FAA/AM-01/06). Washington, DC: FAA Office of Aviation Medicine.

- Waugh, G. (2001). Predictor-criterion analyses. In R.A. Ramos, M.C. Heil, C.A. Manning (Eds.) *Documentation of validity for the AT-SAT computerized test battery, Volume II (DOT/FAA/AM-01/06)* (pp.37-42). Washington, DC: FAA Office of Aviation Medicine.
- Wise, L.L., Tsacoumis, S.T., Waugh, G.W., Putka, D.J., Hom I. (2001). *Revisions of the AT-SAT (DTR-01-58)*. Alexandria, VA: Human Resources Research Organization (HumRRO).

Author Note

Raymond E. King, Carol A. Manning, and Gena K. Drechsler, Aerospace Human Factors Research Division, Civil Aerospace Medical Institute, Oklahoma City, Oklahoma.

The authors gratefully acknowledge Clara Williams for the time she spent building the database that made this report (and the many more that will follow) possible. No research project ever enjoyed the services of a more dedicated professional.

This is a statistical snapshot of the workforce demographics. The use of this data in any employment decision is PROHIBITED without the express written authorization of the Deputy Chief Counsel for Operations, AGC-3.

Correspondence concerning this article should be addressed to Raymond E. King, CAMI/AAM-520, P.O. Box 25082, Oklahoma City, Oklahoma 73125. E-mail: Raymond.King@faa.gov

Footnotes

¹ AT-SAT was approved as the official ATCS selection test for those applicants without previous air traffic control experience on May 13, 2002, with June 2002 marking the first time the test was used operationally.

² Adverse Impact – “A selection rate for any race, sex, or ethnic group which is less than 4/5 (80%) of the rate for the group with the highest rate.” (Uniform Guidelines on Employee Selection Procedures, 1978, Sec 4D.)

A Review of the Current State of Aviation Safety Action Programs in Maintenance Organizations

Manoj S. Patankar

and

Jiao Ma

*Department of Aviation Science
Saint Louis University
3450 Lindell Boulevard
St. Louis, MO 63103
USA*

Abstract

In maintenance, Aviation Safety Action Programs (ASAPs) were designed to encourage air carrier and repair station employees to report their errors or conditions hazardous to safety of flight voluntarily. Such programs have been operational since 1998. This paper presents the results of a survey of 20 organizations, which was conducted to better understand current state of the ASAP programs. A rise in the number of ASAP Memorandum of Understanding and the results of this survey indicate that the ASAP programs are gaining support among an increasing number of maintenance organizations across the country. Successful ASAP programs are found to have strong and consistent support from senior management to the extent that they are able to overcome cases that challenge the corporate disciplinary policy. The resources required to run such programs at different types of maintenance organizations as well as at the local FAA offices are presented. Comprehensive corrective actions resulting from the investigation of ASAP events are reported at three levels of impact: task-level changes, organization-level changes, and industry-level changes. The authors argue that the effectiveness of ASAP programs in maintenance should be measured by the percentage of actual changes at each level rather than the number of total or sole-source reports. As the number of organizations participating in such programs grows, there will be an increased need to share information about intervention strategies and their effectiveness across the participating organizations; therefore, future studies that document a case-by-case analysis of selected ASAP events followed by the development of a common data classification scheme are proposed.

Background

Aviation Safety Action Programs (ASAPs) were designed to encourage aviation professionals to report safety information that may be critical to identifying potential precursors to accidents voluntarily. In 1996, ASAPs were introduced in the flight domain with the hope of encouraging pilots to disclose their errors and the factors contributing to their errors, which were otherwise not likely to come to the attention of the company management or the Federal Aviation Administration (FAA) (Patankar, 2004). The FAA further expanded the scope of this program and added guidance materials for the maintenance, dispatch, and cabin communities (FAA, AC 120-66B, 2002); the first maintenance ASAP program was approved in 1998. As of July 2006, there were 50 flight ASAPs, 27 maintenance and engineering ASAPs, 27 dispatch ASAPs, and 6 flight attendant ASAPs (FAA, 2006). Over the past three years, the numbers of flight and maintenance ASAPs have increased by 17% and 1700% (the initial number was so small; therefore, this number is inflated), respectively. The ratio of flight-to-maintenance programs decreased from about 4:1 to less than 2:1, which had remained steady since early introduction of maintenance ASAPs.

An ASAP program provides for the collection, analysis, and retention of the safety data that is obtained. Safety issues are then resolved through systemic changes rather than through punishment or discipline of individual employees. An ASAP program is based on a safety partnership that includes the FAA, the certificate holder (company), and employees' labor organization representative or a non-union representative -- if the reporting employee is not represented by a labor union (FAA, 2002).

Patankar and Driscoll (2004) conducted a survey of over 5,000 FAA-certificated maintenance personnel, and they discovered two factors that were key to the success of ASAP programs: (a) level of awareness about ASAPs among the maintenance personnel and (b) level of interpersonal trust among employees, company management, and their FAA inspectors. Taylor (2004) noted two other factors that may impact the success of an ASAP program: robustness of the Memorandum of Understanding (MOU)¹ and Event Review Committee (ERC)² process.

In order to assist the maintenance and engineering community in building their ASAPs, the national-level Aviation Rulemaking Committee (ARC) formed a Maintenance and Engineering (M&E) Subcommittee in October 2004. Since then, the M&E Subcommittee has taken a leadership role in providing a coordinated forum for airlines and repair stations to address the unique challenges

¹ MOU refers to the written agreement between two or more parties setting forth the purpose for, and terms of, an ASAP.

² ERC, a group comprised of a representative from each party to an ASAP MOU – a management representative from the certificate holder, a representative from the employee labor association (if applicable), and a specifically qualified FAA inspector from the certificate holding district office. The group reviews and analyzes reports submitted under an ASAP. The ERC may share and exchange information and identify actual or potential safety problems from the information contained in the reports.

and opportunities associated with ASAPs in maintenance. With an expanding membership, its agenda includes raising the awareness of maintenance-specific ASAP issues as well as building an error classification scheme to support the overall industry's Voluntary Aviation Safety Information-Sharing Process (VASIP) (Patankar & Gomez, 2005). The M&E Subcommittee held two information-sharing meetings in 2005 to solicit feedback from maintenance organizations based on their experiences in either trying to get an ASAP approved or in running an existing program. The first meeting was held in June 2005 in St. Louis, MO, and included 30 individuals representing airlines, repair stations, labor unions, FAA Certificate Management Offices, and FAA Headquarters. This group was asked to describe the specific barriers they faced in implementing ASAPs as well as to share some of their success stories. A second meeting was held in October 2005 in Indianapolis, IN, and included 40 individuals representing a similar background as at the previous meeting. This meeting followed-up on issues that were not resolved previously and assessed the status of ASAP programs that were currently in operation or in the process of being accepted. In both information sharing meetings, participants discussed the challenges faced by company managers, labor union representatives, and the FAA inspectors in supporting safety programs and collected recommendations for data analyses and trending.

The key questions that emerged from the information sharing meetings were:

1. What is the current state of the ASAP programs in aviation maintenance-how many reports are being collected, what percentage of those reports are sole-source³ reports, what personnel resources are dedicated to ASAP programs, and what are the typical outcomes of ASAP investigations?
2. What are some of the best practices, especially in increasing the awareness of ASAP programs, used by companies that have successful programs?
3. What are the critical challenges to instituting such programs in additional organizations, including repair stations as well as in strengthening the existing programs?

Method

A 36-item survey instrument (see Appendix) was constructed by the M&E Subcommittee to seek answers to the key questions raised in the information sharing meetings. These items addressed four aspects of an ASAP program: (a) background and structure, (b) current operations, (c) its outcomes, and (d) specific examples of common safety threatening behaviors. The survey contained multiple choice, yes/no, and open-ended questions and took approximately 30 minutes to complete. Survey responses were anonymous, although respondents were asked to indicate their type of employer, and whether they were answering from the point of view of an ERC member who represented management, an employee group or trade union, or the FAA.

³For the purposes of ASAP, a report is considered as sole-source when all evidence of the event available to the FAA outside of ASAP is discovered by or otherwise predicated on the ASAP report. It is possible to have more than one sole-source report for the same event.

In January 2006, 35 questionnaires were sent to all the companies with an approved maintenance ASAP MOU as well as those members of the M&E community who did not have an approved program, but had attended the Maintenance ASAP Information Sharing meetings in 2005. Thus, the population consisted of all individuals who had ample knowledge and direct experience administering maintenance ASAP programs of their own, or who were developing such a program.

The total population of air carriers and maintenance operators with Maintenance ASAP agreements was 27 at the time this survey was administered. Eighteen (66.7%) of them, including the six earliest and most experienced ASAP programs, participated in this survey. The overall actual sample size was 20 (including two companies without a maintenance ASAP).

The survey responses were anonymous, although respondents were asked to indicate the type of their employer and whether they were answering from the point of view of an ERC member who represented management, an employee group or trade union, or the regulator. Four of the twenty valid questionnaires collected were FAA responses, and the rest were multiple organization types, as listed in Table 1.

Table 1
Type of organizations

Type of Organization	Number of Organizations	Average Number of Employees Covered under M-ASAP MOU
Legacy Carrier	7	6,486
Low-Cost Carrier	3	1,050
Regional Carrier	4	2,633*
Repair Station	2	3,344
FAA	4	

Note. *2 out of the 4 regional carriers did not provide these data.

Results and Discussion

This section presents the results of the survey in terms of responses to the key questions raised at the two information-sharing meetings described in the background section.

The Current State of the ASAP Programs in Aviation Maintenance

Based on the data reported by 10 organizations (three legacy carriers, three low-cost carriers, three regional carriers, and one repair station), an average of 146 reports were received in the year 2005. The legacy carriers represented the most employees (6,486) and received the most reports (average of 196); whereas, the low-cost carriers represented the least employees (1,050) and received the least reports (average of 99). Thus, the response supported a positive correlation between the number of employees or the size of the organization and the number

of reports received -- larger the organization, the more reports it is likely to receive.

Maintenance ASAP programs had a consistently high acceptance rate across different organizations (mean = 98.7%, SD = .99); however, percentage of sole-source reports varied considerably among organizations (mean =83.5%, SD = 21.2). For example, three legacy carriers reported that about 69% of their reports were sole-source, and two regional carriers reported their sole-source reports to be at 74.8%; whereas, one repair station reported that 99.4% of their reports were sole-source. Based on the discussions at the information sharing meetings, the guidance in the ASAP Advisory Circular (AC 120-66B), and the specific language used in the respective approved MOU, there are two ways of defining sole-source reports: (a) some organizations define sole-source reports that contain information which is not available to the FAA from any other means; while (b) others define sole-source reports as those containing information not available to the FAA or the Company from any other means. This has given rise to the distinction between the two definitions as sole-source to the FAA and sole-source to the Company, respectively. Clearly, most organizations with maintenance ASAP programs are using the former definition since errors in maintenance are often discovered by a person other than the one who committed the error. Such errors could lay dormant for several months.

Organizations that are interested in starting an ASAP program have often asked, at the information sharing meetings, about the staffing needs for such a program. Technically, two people are required per participating party (i.e., the company, the labor union or employee association, and the FAA Certificate Management Office). One person from each group acts as the primary and voting member and the other serves as the backup person. Additionally, resources may be required to collect data, analyze data, and prepare periodic reports. The 20 people who participated in this survey spent on an average 55% of their time on ASAP responsibilities, with only five people assigned to the job on a full-time basis. On an average, 2.6 full-time positions (combined company and labor) per 100 ASAP reports per year were found to be necessary.

Comprehensive corrective actions resulting from the investigation of ASAP events are reported at three levels of impact: (a) task-level changes, (b) organization-level changes, and (c) industry-level changes (see Figure 1). Typical outcomes of ASAP investigations include changes to the primary maintenance procedures (for example, reorganizing the sequence of tasks so that inspections could be carried out at the appropriate time in the maintenance sequence). Corrections of such procedural issues tend to address latent errors (Reason, 1997), which could have resulted in recurrent errors by the maintenance technicians. The researchers classified such outcomes as task-level changes and about 70-75 percent of the changes reported by maintenance ASAP programs fall in this category. As the primary maintenance procedures are being refined, some organizations are starting to address secondary procedures (for example, improving the process used to prepare/procure and quality-check maintenance kits) and organizational policy issues. Solutions to such problems usually affect a larger segment of the organization and hence the researchers classified such outcomes as organization-level changes. About 20-25 percent of the changes reported by

maintenance ASAP programs fall in this category. The next level of outcomes is classified as industry-level changes because they tend to impact multiple organizations (for example, a latent error in published maintenance procedures or recommended repair procedures may impact a certain aircraft model across multiple organizations). These changes account for 1-5 percent of total maintenance ASAP reports.

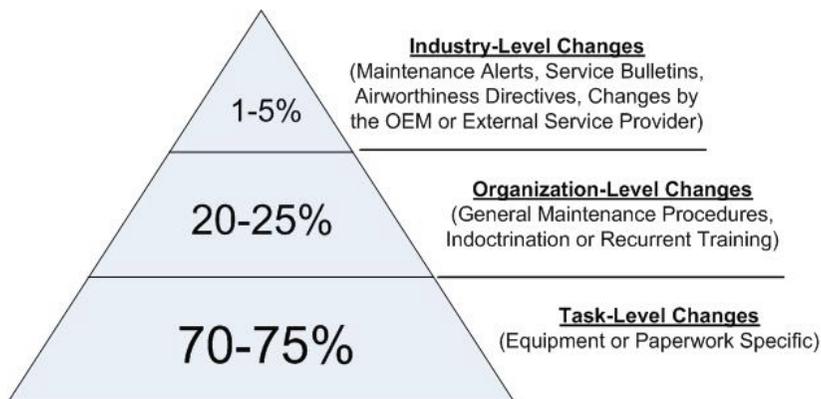


Figure 1. Changes driven by ASAP events display a pyramid-like relationship.

The limited volume of maintenance ASAP reports (as compared to flight ASAP reports), the complexity of maintenance investigations (Patankar & Driscoll, 2004; Patankar & Gomez, 2005), and the nature of outcomes (e.g., latent maintenance errors, Hobbs, 2004) resulting from maintenance ASAP investigation, collectively strengthen the characterization of the maintenance environment as a network of different organizational units potentially impacting the quality of maintenance.

For example, upon receiving an ASAP report, the ERC has to validate and investigate every piece of data in that report. Such process could lead to (a) documents such as maintenance manuals and publications, logbooks, training content and records, (b) external vendors or contract maintenance, policies, procedures, and (c) parts availability and equipment condition. Once all the facts are collected and reviewed, the ERC needs to decide, unanimously, if this report is acceptable and corrective actions, in accordance with the ASAP MOU.

ASAP investigations of most maintenance errors lead to the discovery of either task-level changes or organization-level changes that should be addressed by appropriate comprehensive resolutions. Therefore, the researchers believe that effectiveness of ASAP programs in maintenance should be measured by accounting for what has actually changed in the organization and the percentage distribution of actual changes at each level (task, organization, or industry) rather than the number of sole-source or non-sole source reports. The researchers posit that the distribution of the corrective actions ratios across the task-level, organization-level, and industry-level changes will change as the specific ASAP program matures. For instance, data from surveys, information sharing meetings, and field visits indicate that at the early stages of the ASAP programs, most of the

changes tend to be concentrated at the task level. As the program matures, they tend to discover more systemic errors and consequently address organizational issues. Once the task-level and organization-level issues are stabilized, the program seems to be better equipped to handle industry-level challenges. This does not rule out the possibility that new ASAP programs will make an industry-level contribution.

Best Practices in Increasing Awareness

All organizations used multiple channels to educate/inform their workforce about the ASAP program prior to its implementation (see Figure 2). The most common one was face-to-face small group meeting, which was used by all the organizations. Additional methods reported include computer-based training, recurrent training, road show, and brochure.

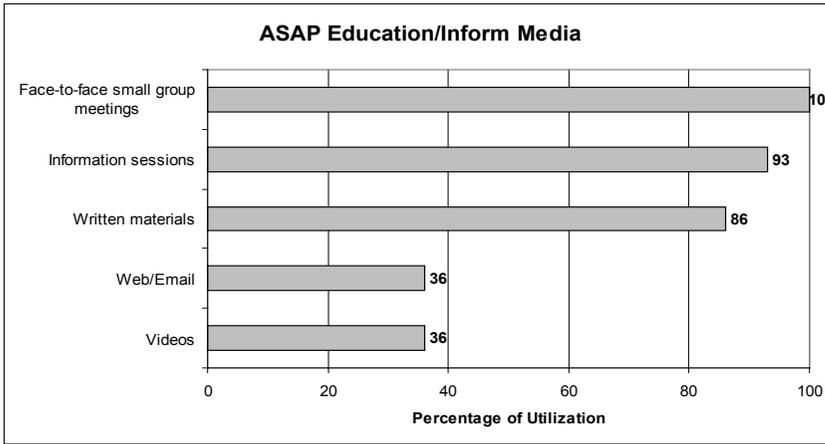


Figure 2. Common education/inform media for ASAP programs.

Companies with successful ASAP programs indicated that their management was very supportive of the ASAP programs, and their commitment appeared in providing resources for the ERC, periodically reviewing reports on ASAP activities, speaking highly of the program, and ensuring that the middle managers take ASAP ERC's recommendations seriously.

There are multiple means to leverage the success of a maintenance ASAP program. Some companies reported the use of ASAP results in their Continuing Airworthiness Surveillance System (CASS) reports; strong articulation of the ASAP findings with the ongoing human factors training program; and periodically discussing the state of the ASAP programs across multiple professional groups - flight, dispatch, and flight attendant.

Critical Challenges Ahead

First, commitment by senior management (managing director and above) is critical to the success of ASAP programs because such programs require all the managers in the typical chain of command to focus on the systemic improvements that could be made rather than punishing the individual who committed the error. This is a significant fundamental shift in attitude and behavior for many

managers and it tends to challenge the corporate disciplinary policy. The current survey indicates that only 53% of the organizations consider the support from their managers to be consistent throughout their organization, and senior managers rarely attended ERC meetings. In order to improve the overall strength of management's commitment to ASAP programs, the researchers strongly recommend actively inviting a wide range of line managers to ERC meetings and including them in the ASAP feedback/communication loop.

The survey responses revealed that organizations with successful ASAP programs have chosen not to exercise their disciplinary policies until the ASAP investigation has been completed. They have consistently yielded in favor of the ASAP program. In the case of non-sole-source reports, once a report was accepted by the ERC, 73% of the ASAP MOUs did not allow for disciplinary action by the company against the individual. None of the companies planned to include such a provision in its MOU in the future because a potential corporate disciplinary action might very likely result in loss of participation for established ASAPs and hinder participation for upcoming programs. For those companies whose ASAP MOUs allowed corporate disciplinary action noted that in practice, it was rarely used because of the possible negative effect that such an action might have on the overall success of the ASAP program.

Second, continuing to raise the awareness regarding ASAP programs, especially the effect such programs have had on the overall safety culture at each participating organization is critical to sustaining the current momentum. The first three maintenance ASAP programs were approved in 1998; the next three programs were approved in 2001; and the next four programs were approved in 2004. During those times, the major barrier was that of awareness regarding the nature of the ASAP program, its value to the participating organizations as well as individuals, and the level of protection to the individual. Now that there are 27 maintenance ASAP programs, the need seems to have shifted from building general awareness toward communicating the actual changes resulting from ASAP programs. Figure 3 illustrates the means that are currently used to measure the success of maintenance ASAP programs.

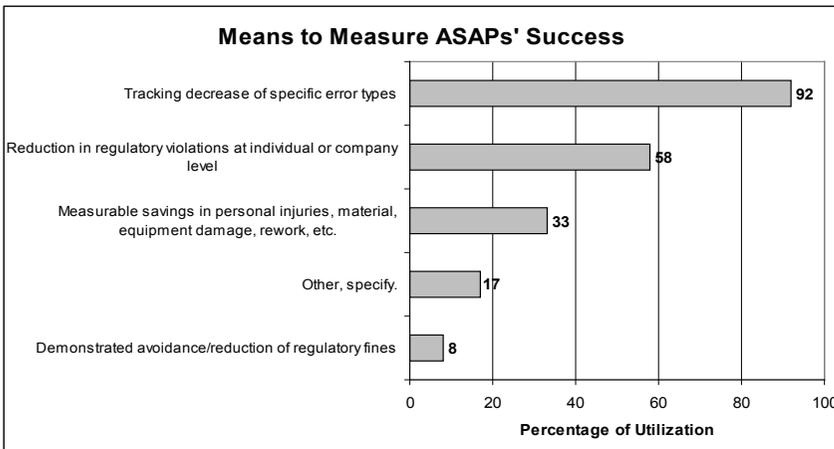


Figure 3. Means to measure ASAP programs' success.

Despite the organization's size, the ERC at any organization is relatively limited in resources and in its capacity to track and identify common contributing factors across industry and trends of intervention strategies and their outcomes over time. As the aviation industry moves toward aggregation of safety data such that pertinent, but de-identified information may be shared across multiple ASAP programs (flight, maintenance, dispatch, and flight attendant) in the same company or multiple ASAP programs across multiple companies, there is a strong need for a standardized data classification scheme for both the initial data (such as event type, error type, contributing causes) as well as the corrective action data. Based on the responses to this survey, most organizations reported using an in-house system to classify their ASAP reports. Classifying mechanism includes revised Maintenance Error Decision Aid (MEDA) forms, Air Transport Association (ATA) codes, risk matrix, and self-defined categories. Some of the reports were categorized on a case-by-case basis through discussions among ERC members.

Third, most (80%) of the FAA Part 121 operators with ASAPs did not include Part 145 operators (external repair stations -- contract maintenance providers) in their MOUs. There are several business, legal, and logistical issues that inhibit participation of repair stations at this time; however, their continued participation in the ASAP information sharing meetings indicates that there is strong interest in at least some of the major repair stations. Efforts need to be made to build alliances such that participation in the ASAP programs is viable for such repair stations. A model similar to the one used by the Medallion Foundation in Alaska (Medallion, 2006) has been suggested at the June 2006 information sharing meeting.

General Discussion

Overall, the challenges of maintenance ASAP programs have expanded from the early need to increase awareness to the current need to increase the ability of such programs to change the organizational safety culture and measurement of these changes. Efforts are underway to develop a standardized data classification scheme so that maintenance errors may be classified and compared across multiple organizations as well as errors and contributing factors across multiple ASAP programs within the same organization. Such efforts are expected to exponentially increase the impact of ASAP programs and thereby have a profound impact on the overall improvement of the safety culture at the organization as well as industry level.

It has been eight years since inception of the first maintenance ASAP, and the number of maintenance ASAP programs has quadrupled over the past three years. The researchers believe that evaluation of ASAP programs using consistent and multi-dimensional (e.g., error reduction, cost saving) measures could enable the program managers to establish compounding effects of systemic changes resulting from an ASAP investigation. It is hypothesized that the higher a specific change initiative in the Figure 1 hierarchy is, the greater the impact of that change -- industry-level changes will have the greatest impact.

Like other safety programs, maintenance ASAP programs can be measured by their impact on safety operations both inside a specific organization and

industry wide, and safety data can be converted to or affiliated with monetary value in order to justify their return on investment (ROI). For instance, common results of maintenance errors like flight cancellations and in-flight shutdowns can be assigned a dollar value and the financial impact of maintenance errors that are reported through ASAP reports could be quantified. Once the average cost of these events, by fleet type, is known, it would be powerful to present the impact of reductions in maintenance errors.

Conclusions

Overall, the number of Maintenance ASAP programs has grown from 6 to 27 in the last three years. In parallel with this growth, the maintenance community has developed an ARC endorsed M&E Subcommittee that has held four Information Sharing meetings and conducted a survey to assess the current state of the ASAP programs. The following conclusions could be drawn from the data presented in this article:

1. Successful ASAP programs tend to have strong and consistent support from senior management to the extent that they are able overcome cases that challenge the corporate disciplinary policy.
2. Impact of ASAP investigations could be classified into three categories: task-level, organization-level, or industry-level. Currently, most of the changes resulting from ASAP investigations tend to impact at the task level. In some cases (approximately 1-5% of the total cases), the impact has been at the industry level.
3. There are several ways to measure the effectiveness of ASAP programs; emphasis on the actual changes accomplished as a result of the ASAP investigations, rather than only the volume of reports, would be more useful in assessing the program effectiveness.
4. As the number of ASAP programs continues to grow, there will be an increased need to share information about intervention strategies and their effectiveness across the participating organizations. Future research and development efforts need focus on building standardized data classification schemes for both event and contributing factors as well as for corrective actions.

Acknowledgements

This research was supported by FAA Grant #2003-G-013. Special thanks to Mr. Rick Yorman from American Airlines and the M&E Subcommittee, and Dr. Thomas Longridge, Dr. William Johnson, Dr. William Krebs, and Mr. John J. Hiles from the FAA for their continued support and encouragement. On behalf of the maintenance and engineering community, we also thank the ASAP/FOQA ARC for their support to the M&E Subcommittee. Of course, this study would not have been possible without the participants in the survey -- we are deeply grateful for their contributions.

References

Federal Aviation Administration (2002). Advisory circular 120-66B: *Aviation safety action program (ASAP)*. Retrieved November 1, 2006, from <http://www.air->

web.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/0/61c319d7a04907a886256c7900648358/\$FILE/AC120-66B.pdf

- Federal Aviation Administration (2006). *Aviation safety action program topics – participants*. Retrieved July 12, 2006, from http://www.faa.gov/safety/programs_initiatives/aircraft_aviation/asap/participants/
- Hobbs, A. (2004) Latent failures in the hangar: uncovering organizational deficiencies in maintenance operations. *International Society of Air Safety Investigators Annual Conference, 2004*. Gold Coast, Australia, August 29 – September 2, 2004.
- Medallion Foundation. (2006). *Medallion foundation safety star process*. Retrieved August 8, 2006, from <http://www.medallionfoundation.org/documents/MedallionFoundationSafetyStarProcess.pdf>
- Patankar, M. S. (2004). *Development of guidelines and tools for effective implementation of an aviation safety action program (ASAP) for aircraft maintenance organizations: First year report*. FAA Grant Number 2003-G-013. Retrieved March 1, 2006, from <http://www.hf.faa.gov/docs/508/docs/MaintASAPYr1.pdf>
- Patankar, M., & Driscoll, D. (2004). *Factors affecting the success or failure of aviation safety action programs in aviation maintenance organizations: Second year report*. FAA Grant Number 2003-G-013. Retrieved March 1, 2006, from <http://www.hf.faa.gov/docs/508/docs/maintFY04ASAPrpt.pdf>
- Patankar, M., & Gomez, M. (2005). *Maintenance ASAP programs: Barriers and opportunities*. Retrieved March 1, 2006, from <http://www.hf.faa.gov/docs/508/docs/Av-Maint05.pdf>, pp. 35-42.
- Reason, J. (1997). *Managing the risks of organisational accidents*. Aldershot, UK: Ashgate Aviation.
- Taylor, J. C. (2004). *Prototype training materials for acceptance criteria of maintenance ASAP events occurring within social context*. Retrieved May3, 2005, from http://www.faa.gov/safety/programs_initiatives/aircraft_aviation/asap/reports_presentations/media/MaintenanceASAP.pdf

Appendix: Maintenance ASAP Survey Instrument

January 2006

Maintenance ASAP Survey

1. Select your employer type:
 - Legacy Carrier
 - Low Cost Carrier (LCC)
 - Regional
 - FAA
 - Repair Station
 - Other (please specify) _____

2. Select the type of ASAP program(s) currently in operation at your organization (check all that apply).
 - Maintenance
 - Flight
 - Dispatch
 - Flight Attendant

3. When was your Maintenance ASAP program FIRST approved?
_____ (Mo, Year)

4. Are your mechanics represented by a labor union?
 - Yes
 - No

5. How does your ASAP program interact with your human factors program (CRM/MRM or Maintenance Human Factors)? (E.g.: CRM/MRM training discusses ASAP process)

6. Does your ASAP MOU allow non-FAA certificated employees to submit ASAP reports?
 - Yes
 - No*If YES, please specify the employee groups (E.g.: Engineering, GSE, Stores, Fuelers, etc.)*

7. If non-FAA certificated employees are allowed to submit ASAP reports, how are such groups represented on the Event Review Committee?

8. Which of the following categories do you personally represent on the ERC or the ASAP program?
 - Company
 - Union
 - Employee Group
 - FAA
 - Other, specify: _____

9. What percentage of your time/job is assigned to ASAP responsibilities?

10. Do you have any other ASAP programs that will have a signed MOU within the next year?
 - Yes
 - No

11. How many people are assigned to your ASAP group (fulltime equivalent)?
 - Company: _____

- Union: _____
 - Employee Group: _____
 - FAA: _____
 - Other (analysts, etc.): _____
12. How many changes have been effected/driven by ASAP events in each of the following categories?
- Procedures (Equipment or paperwork specific): _____
 - Policy (Company-wide/GMM changes): _____
 - Industry-wide (ADs, Service Bulletins, etc.): _____
13. Does your program allow you to inspect an aircraft without delay or red-tape?
- Yes
 - No
- If NO, what holds the process up? _____
14. Do you think that sharing ASAP data with NASA's ASRS is effective and adds value to the ASAP program?
- Yes
 - No
- If YES, why should both programs be used? _____
- If NO, explain why. _____
15. Please provide the following data regarding your maintenance ASAP submissions for 2005.
- Total number of reports received: _____
 - Number of Reports Accepted: _____
 - Number of Reports Rejected: _____
 - Number of Sole-source Reports: _____
 - Number of Non-sole-source Reports: _____
16. Please state the number of employees covered under your maintenance ASAP MOU.

17. In your opinion, what is the general reaction of the employees regarding ASAP submissions that are ACCEPTED by the ERC?
- Satisfied
 - Neutral
 - Not Satisfied
18. In your opinion, what is the general reaction of the employees regarding ASAP submissions that are REJECTED by the ERC?
- Satisfied
 - Neutral
 - Not Satisfied
19. How did your company educate/inform the workforce regarding the ASAP program prior to its implementation? (Check all that apply)
- Videos
 - Web/E-mail
 - Information sessions
 - Written materials
 - Face-to-face small group meetings
 - Other, specify: _____

20. How does your company measure success of the ASAP program? (Check all that apply)
- Tracking decrease of specific error types
 - Reduction in regulatory violations at individual or company level
 - Demonstrated avoidance/reduction of regulatory fines
 - Measurable savings in personal injuries, material, equipment damage, rework, etc.
 - Other, specify: _____
21. How is your maintenance ASAP program included in your Continuing Analysis and Surveillance System (CASS)? (Check all that apply).
- QA Audits
 - Logbook pages for signoffs
 - Other, specify: _____
 - None of the above
22. Does your ASAP program allow for submission of non-airworthiness events or safety/hazard submissions (e.g., safety concerns)?
- Yes
 - No
23. For events discovered by the company (non-sole-source events), how are the affected employees notified of the event?
- Via the union or employee group representative
 - Via the local management
 - Via the company's safety
 - Via the ERC member
24. Please explain how your senior management (managing director and above) supports the ASAP program?
- _____
- _____
- _____
25. How often does senior management attend ASAP meetings?
- All the time
 - Sometimes
 - Once in a while
 - Never
26. If the senior management does not attend, how are they notified?
- Verbally
 - Via E-mail
 - Other, specify: _____
27. Indicate your level of agreement with the following statement: Management support of the ASAP program is consistent throughout our company.
- Strongly disagree
 - Disagree
 - Neutral
 - Agree
 - Strongly agree
28. Does your ASAP MOU allow for disciplinary action by the company against the individual (in the case of non-sole-source reports), if their report is accepted?
- Yes
 - No
- If NO, are you planning on including such provision in your MOU in the future?*

- Yes
- No

29. For approved or established programs: If your ASAP MOU were changed to allow the company to take disciplinary action (in the case of non-sole-source reports), to what extent are you likely to lose participation?
- Completely
 - Somewhat
 - Not at all
30. For companies without ASAP programs: If your ASAP MOU were to allow the company to take disciplinary action (in the case of non-sole-source reports), to what extent are you likely to risk gain participation?
- Completely
 - Somewhat
 - Not at all
31. For 121 Operators with ASAP programs: Does your MOU cover partner 145 operators (primary maintenance providers)?
- Yes
 - No

32. Suggest ideas to improve participation by 145 providers:

33. How do you classify your ASAP reports? (E.g. do you use the MEDA form or do you have your own system?)

34. How do you classify the corrective actions to specific ASAP events? If you don't classify corrective actions, please state so.

35. Do you prepare comprehensive reports for each ASAP submission—similar to an accident investigation report?
- Yes
 - No

36. Please provide examples of the following types of behaviors:
- At-risk behavior:

 - Negligence:

 - Reckless behavior:

Flight Crew Callouts and Aircraft Automation Modes An Observational Study of Task Shedding

Örjan Goteman

*Lund University
Department of Design Sciences
Box S-255 91 Lund, Sweden
tel: +46 42 911 17
orjan.goteman@webbpost.nu*

and

Sidney Dekker

*Lund University
School of Aviation
SE-260 70 Ljungbyhed, Sweden
Sidney.dekker@tfhs.lu.se*

Abstract

New aircraft come with a set of recommended standard operating procedures, in the case of multi-crew aircraft this includes “callouts”—verbalizations of particular flight guidance automation mode changes. In an attempt to reduce the risk for mode confusion some operators have required flight crews to callout all flight guidance automation mode changes as a means of forcing pilots to monitor the Flight Mode Annunciator (FMA). Previous research has shown that crews do not spend enough time on the flight mode annunciator, and skip mode call-outs as well as making call-outs in advance of annunciations; there has been no report of any system or regularity in the shedding and adaptation of callouts. One reason could be the contrived empirical simulator settings of such research, which we aimed to augment with natural observations of real cockpit work reported here. With the hope of answering, in more detail, how required verbal coordination of annunciated mode changes gets adapted to real settings we observed 19 line flights with three different airlines from the first observer’s seat in the cockpit. We found that many callouts were simply shed in high-workload situations, and found regularity in the kind of callouts being shed. Callouts relating to aircraft automation, such as FMA call-outs, were shed before other required callouts. Our results suggested that FMA callouts were not used as a tool to detect or remember automation mode changes but as a vehicle for coordinating between the pilots themselves, a finding that could serve as a reminder for future design of callout procedures.

Introduction

The aviation industry considers Standard Operating Procedures (SOP's) the backbone of safe operations (JAA 1997, ICAO 2001) and flight crew verbal "call-outs" form a part of these recommended standard operating procedures (Airbus Industrie, 2006, Boeing Commercial Aeroplanes, 1999). "Call-outs" are what one crewmember has to say to the other(s) in a particular operational situation and are intended to ensure effective crew communication, promote situational awareness, and ensure crew understanding of systems and their use (Airbus Industrie, 2006). In an attempt to reduce the risk for mode confusion (see e.g. Sarter & Woods, 1997) some aircraft manufacturers have required flight crews to callout all flight guidance automation mode changes as a means to force pilots into monitoring the flight mode annunciator. The underlying strategy is that there is a need for pilots to know the actual flight guidance mode at all times and that by requiring all mode changes be called out, the pilots will spend more time on the flight mode annunciator (FMA) and, presumably, their mode awareness will increase (Airbus Industrie, n.d.).

Though pilots need to know the flight guidance mode at all times, they do not dwell much on the flight mode annunciator. The average cumulative dwell time on the flight mode annunciator is as low as 2.9% of the time (Hüettig & Anders, 1999). In addition to the low dwell time, detection rate for unexpected mode changes is low—in the best cases 60% (Mumaw, Sarter & Wickens, 2001). Unexpected mode changes have been implicated in accidents with automated aircraft (e.g. FAA, 1996; Sarter & Woods, 1997), which raises questions about pilots' abilities to keep track of automation. Does this assumption work in the operational reality? Previous research confirmed how pilots (as do other operators) adapt the application of procedures to practical task demands (Snook, 2000; Dekker, 2003) and this applies to mode callouts too (Degani & Wiener, 1994; Huettig, Anders & Tautz, 1999; Plat & Amalberti, 2000; Mumaw et al., 2001; Björklund, Alfredsson & Dekker, 2006). Yet the latter work revealed little regularity in which callouts are adapted, or how; therefore, it has limited leverage over how to potentially intervene—procedurally or through design—to improve mode awareness. One reason for this limitation could be that this work has mostly been carried out in simulated settings, studying the eye movements or other parameters of a single pilot in the studied flight crews (Björklund et al., 2006 being an exception to the latter). Recent applications of discourse analysis in aviation human factors (e.g., Nevile, 2004) encouraged the study of talk-in-interaction in natural settings, in order to go beyond stylized, partial, or static descriptions of the work as it occurs naturally—with two pilots jointly having to monitor and make sense of the behavior of their automation. The notion of talk-in-interaction deliberately suggests "talk is not all participants do as they interact" (Nevile, 2004, p. 21), which leaves analytic room for other resources to be drawn on (e.g., the FMA, pointing) as participants in the work jointly form meaning around the setting they interact in. The purpose of this paper is to report on an effort to augment current knowledge on mode monitoring with such a talk-in-interaction study. We hope to answer, in more detail, how verbal coordination of annunciated mode changes gets adapted in real settings.

As with previous research on unexpected mode changes (Sarter & Woods, 1997), the majority of the observed flights in this study were performed on the

A320 family aircraft. We use the Airbus Industrie's term Flight Management Guidance System (FMGS) to denote the whole flight guidance system. The FMGS hardware artefacts visible to the pilots are the Flight Mode Annunciator (FMA), the Flight Control Unit (FCU) and the Multipurpose Control and Display Unit (MCDU).

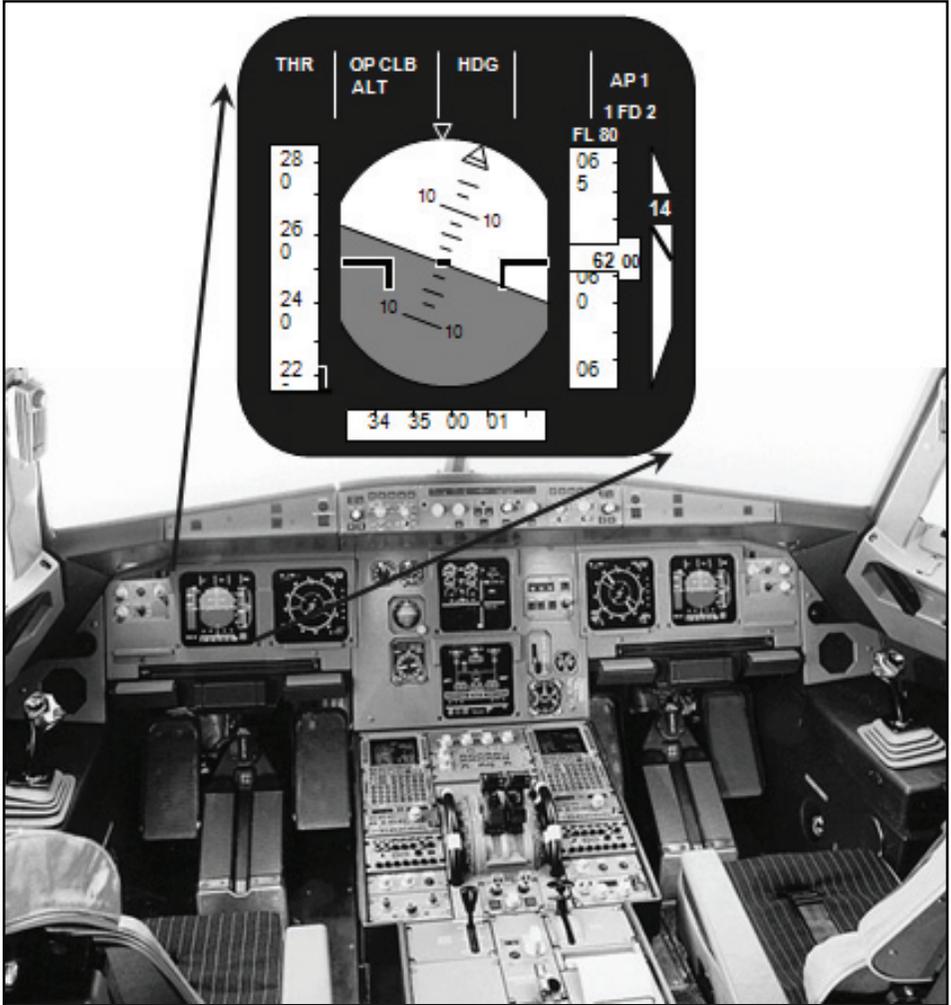


Figure 1. The Flightdeck of A320 with the FCU on the glare shield and the Primary Flight Display with the FMA above the artificial horizon. The left pilot's Primary Flight Display is magnified to show the FMA.

The purpose of the FMGS is to aid the pilot in achieving the objectives of flying the aircraft safely and efficiently from takeoff up to and including landing and freeing the pilots from hand flying the aircraft. The FMGS is capable of automatically changing flight guidance modes as it becomes necessary for the flight guidance system to follow the pre-planned flight plan. The pilot can also select modes and target values through the Flight Control Unit mounted on the glare

shield panel in front of him or her. Aircraft produced after 1980 typically display the active and armed automatic flight guidance modes on a Flight Mode Annunciator (FMA). The FMA is situated on top of the primary flight display, just above the artificial horizon. See Figure 1. The horizon is the “anchor” of the pilots’ instruments—enjoying an average dwell time of 40% of total scanning (Anders, 2001; Mumaw, Sarter & Wickens, 2001). Flight guidance mode status is shown on the FMA as contractions of the mode name in capitals, e.g. “HDG” for “heading select” mode and “CLB” for “climb” mode. Upon a mode change, a frame is shown around the annunciation for 10 seconds.

The viewing angle from the design pilot eye point between the centre of the artificial horizon and the FMA is approximately nine degrees. As this is more than the three degrees angle of focal vision, FMA monitoring must rely on deliberate scanning strategies, and indeed, on cockpit coordinative work other than mere looking. Such action includes the verbal announcements of mode changes, or pointing or nodding to various displays that represent something of apparent interest. Awareness of mode status then, is (or should be) collaboratively produced drawing on multiple sources. This activity, however, is itself embedded in an environment with many concurrent task demands, where it is unlikely that procedures can always be completed from top to bottom or applied linearly, and where pilots do not possess full control over their execution (Loukopoulos, Dismukes, & Barshi, 2003). Naturally occurring ebbs and flows in task load mean that some procedures or tasks will be deferred, interleaved, or shed altogether, and mode callouts are no exception. The questions for the research in this paper included: what mode callouts are typically shed? Are there particular phases of flight more vulnerable to callout shedding? Does callout shedding vary with pilot experience on type?

Method

The Participants

We were able to study how crews worked to coordinate their actions with the flight guidance system on three European air carriers flying the Airbus A320 series aircraft. The flights were randomly selected from each operator’s time table and the commander of the selected flight was contacted before the flight to obtain the pilots’ consent.

The Operational Environment

All flights were performed in Europe in an area of medium to high traffic density. The supporting navigation and air traffic control infrastructure were of high quality. No extreme weather phenomena except cold weather operations with icing and contamination on runways interfered with the operations.

The Observations

Three European air operators were followed over a period of two months. One of the authors acted as observer. The observer was a subject matter expert intimately familiar with the A320 family aircraft and its flight guidance systems (with no active role in the A320 fleet of that airline). He was an active pilot with 6000 hours experience from commercial jet operations, holding a valid B737-NG type certificate. He also had experience from aircraft evaluation and specification

work for an airline, including A320, A330, and A340 aircraft specifications. During the flight, the observations targeted flight guidance mode annunciations, perceived task load, and flight crew callouts. High taskload periods in this study were defined as periods when the observer judged the pilots to be so occupied by their present task that it prevented them from performing additional tasks. The observations were noted by hand by the observer seated in the first cockpit observer seat and later processed for analysis. A total of 19 flights on the A320 were observed.

Participating Operators and Callout Procedures

Prescribed callout procedures differed between the three participating airlines. Operator A required the pilots to callout all FMA changes as they occurred. Operator B required no flight guidance automation callouts at all, whereas operator C required a subset of mode changes to be called out by the pilot flying. This subset included all verified selections (made by the pilots themselves) in addition to modes that altered the level of aircraft automation (for example from a more strategic navigation (NAV) mode to a more tactical heading select (HDG SEL) mode).

Justification for these different callout philosophies was different. Operator A wanted to force their pilots to know and monitor the FMA at all times, while operator B emphasized the importance of raw data showing the actual progress of the aircraft. Operator C justified requiring the callout of only a subset of annunciated modes because it wanted crews to anticipate what the flight guidance automation would do in the near future (see also Sarter & Woods, 1997).

A Normal Flight

For purposes of analysis, a normal baseline flight can be used to lay out the numbers and kinds of mode changes that can be expected. A flight can be divided into various flight phases: Preflight, Takeoff, Climb, Cruise, Descend, Approach, and Landing. A typical line flight can be expected to follow this normal phase progression, a fact utilized by the flight guidance system to set different target speeds for different phases of flight. Such a normal phase progression will also lead to a predictable pattern of flight guidance system mode transitions during the flight.

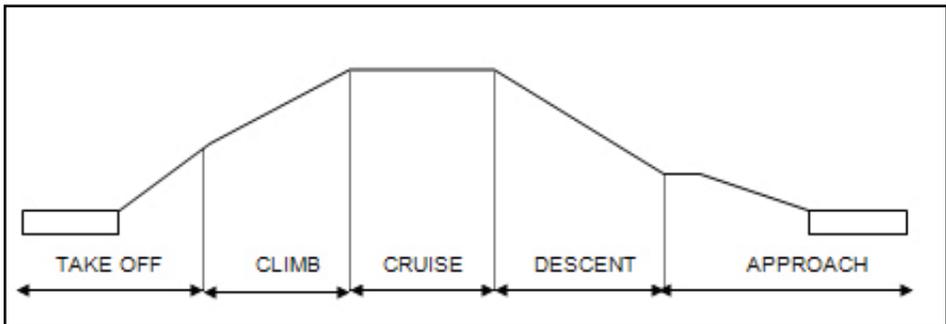


Figure 2. Flight phases of a normal flight in the A320 family aircraft

In Figure 2, we have plotted a non-complicated normal flight. The normal flight actually differs in typical ways from the canonical one predicted by the man-

ufacturer (and on which the flight guidance system logic as well as much training is predicated). The greatest difference is in the use of lower-automation tactical modes (e.g. heading select) in the climb and approach phases in response to air traffic control clearances. Shifts to tactical lateral modes entail automatic shifts in vertical mode. At each flight phase shift, there are a number of typical mode transitions. In addition, there are also some mode transitions typically occurring within the flight phase. In addition, coupled to the vertical mode transitions, there are also a number of mode transitions of the autothrust system. In Table 1, we have listed the number of mode transitions that typically would occur during the normal flight.

Table 1
Mode Transitions

Flight Phase	Typical flight guidance modes for a normal flight		
	Autothrust	Vertical	Lateral
Preflight	blank	blank	blank
Takeoff	MAN FLX	SRS	RWY
Takeoff	MAN FLX	SRS	NAV
Climb	THR CLB	CLB	NAV
Climb	THR CLB	OP CLB	HDG
Cruise	SPEED	ALT*	NAV
Cruise	SPEED	ALT	NAV
Cruise	SPEED	ALT CRZ	NAV
Descend	THR IDLE	DES	NAV
Descend	THR IDLE	OP DES	HDG
Descend	SPEED	ALT*	HDG
Descend	SPEED	ALT	HDG
Descend	THR IDLE	OP DES	HDG
Descend	SPEED	ALT*	HDG
Descend	SPEED	ALT	HDG
Approach	SPEED	ALT	LOC*
Approach	SPEED	ALT	LOC
Approach	SPEED	G/S*	LOC
Approach	SPEED	G/S	LOC
Approach	SPEED	LAND	LAND
Approach	blank	ROLLOUT	ROLLOUT
Sum of mode changes	7	17	9

Depending on the nature of the operational area, the total number of mode transitions occurring over the normal flight may differ. We assumed an uncomplicated flight in a radar environment, including radar vectors for final approach, between two normal, large European airports.

Applying these mode changes to the normal flight flown by the three operators with their different procedures, we expected around 33 flight guidance callouts for operator A, zero for operator B and 14 callouts for operator C.

Results

The data set consisted of 19 flights on the A320 family aircraft with three independent air operators. During these 19 flights, we observed a total number of 589 annunciated mode changes. The majority of the observed flights were with operator C for logistical reasons. See Table 2 for a compilation of the data set.

Table 2
Data Set for Mode Changes

	Operator A	Operator B	Operator C	All
Flights	2	2	15	19
Mode changes	69	56	464	589
Average	34	28	31	31
Median	34	28	32	30

The average number of mode transitions per flight was 31, with the majority (21) occurring during descent and approach phases, see Figure 3 for the average distribution of mode changes during the flight. The median number of mode changes per A320 flight of the study was 30. Of the 589 mode changes 141 were autothrust changes, 307 vertical mode changes and 141 lateral mode changes.

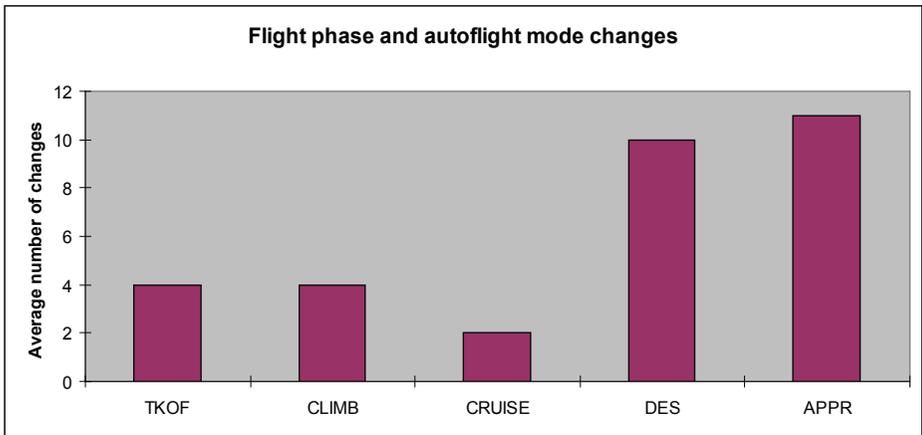


Figure 3. Average number of mode changes per flight phase

Mode Callouts

The average number of mode callouts per flight was over 26 or conversely, 4.6 callouts (15%) per flight were shed. From a total of 87 observed shed callouts, 51 were vertical mode callouts, 27 were lateral mode callouts and 9 were autothrust mode callouts, giving a shedding rate of 17% for vertical mode callouts, 19% for lateral mode callouts and 6% for autothrust mode callouts (see Figure 4).

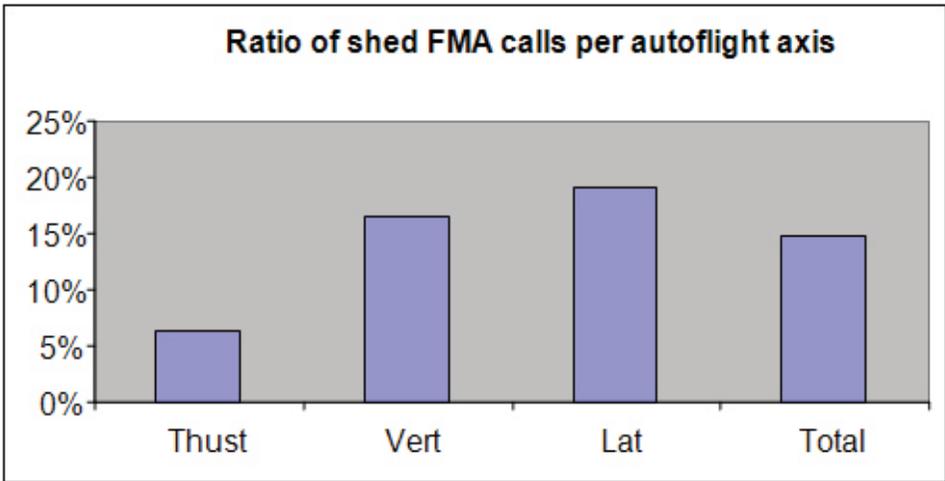


Figure 4. Shed mode callouts per flight guidance axis

There was no significant difference between callout shedding for autopilot vertical and lateral modes, whereas a comparison between autothrust and autopilot mode callouts showed difference in callout shedding, $\chi^2 (df = 1) = 10.36, p < .01$. All callouts, that were called at all, were called out in the prescribed language and form. We noted no improvisations. Callouts that were shed were not recalled or saved for later verbal announcement.

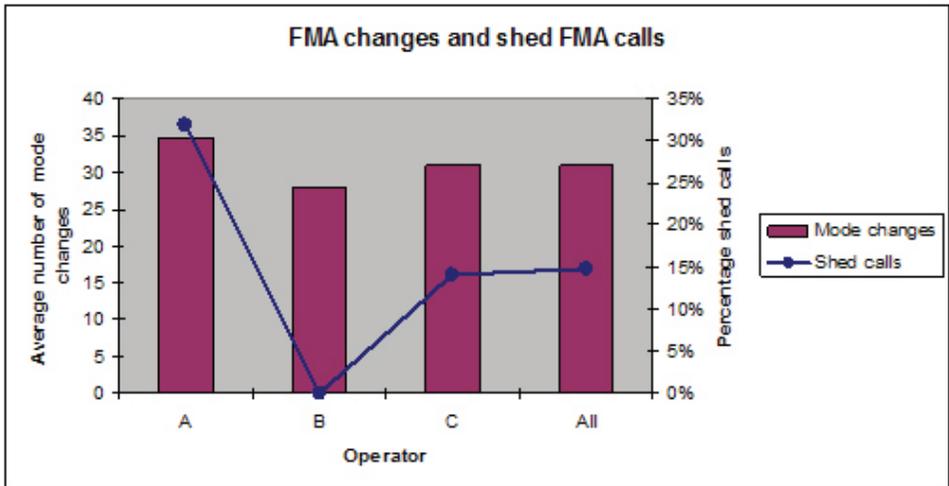


Figure 5. Average number of mode changes and callout shedding per operator

We observed a total of 69 mode changes with operator A, where 22 mode callouts (32%) were shed. We observed 464 mode changes with Operator C, where 65 mode callouts (14%) were shed, see Figure 5. We have cast the figures for operators A and C in a contingency in Table 3 excluding operator B from the comparison as it did not require any mode callouts at all. The differences were significant $\chi^2 (df = 1) = 14.05, p < .001$.

Table 3
Operator A and Operator C Comparison

	Operator A	Operator C	Combined
Mode changes called out	47 (57,7)	399 (388,3)	446
Mode changes not called	22 (11,3)	65 (75,7)	87
Total	69	464	533

In addition to the mode callouts required by procedures, we observed 39 non-required mode callouts at operator B (3) and C (36), and none at operator A. Twenty-three of those callouts were related to the autothrust going to IDLE mode as a consequences of the pilot selecting a descend autopilot mode (mostly in response to air traffic control clearances that required a deviation from the pre-programmed flight path). At operator B, two of the non-required mode callouts were the pilot calling out “Open Climb” mode, which came as a so-called reversion mode when the pilot selected the heading mode on the flight control unit.

Effect of Pilot Experience on Type

The participants had a total average flying experience of 9100 hrs, with an average of 510 hours on the A320 series. The lowest number of hours on the A320 series aircraft of the participants was 10, the highest 5000. The pilot flying had less than 300 hours experience on the Airbus family aircraft in 11 of the 19 observed flights. Those relatively inexperienced Airbus pilots omitted 31 of the totally 87 observed omitted callouts emanating from 589 mode changes. There were no significant differences between the groups, $X^2 (df = 1) = 0.285, p > .59$.

Effect of Flight Phase and Task Load

Mode callout shedding varied between 25% in climb and 11% during approach, see Figure 6.

Collapsing data over flight phase into two groups, one early part and one later part of the flight shows a significant difference between Takeoff to Climb, versus the Descend and Approach phases, $X^2 (df=1) = 10,33, p < .005$. The data are presented in Table 4.

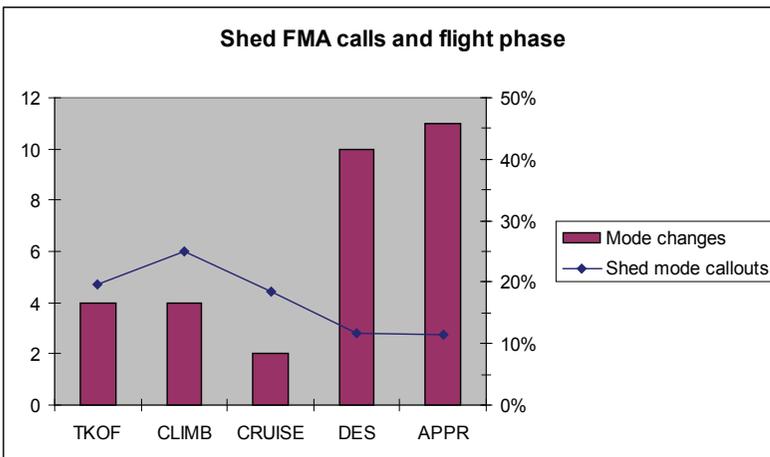


Figure 6. Flight guidance mode changes and callout shedding per flight phase

Table 4

Flight Phase Data

<i>Flight phase</i>	<i>TKOF</i>	<i>CLIMB</i>	<i>CRUISE</i>	<i>DES</i>	<i>APCH</i>	<i>Combined</i>
Mode changes called out	61 (64.8)	57 (64.8)	31 (32.4)	168 (161.9)	185 (178.1)	502
Mode changes not called	15 (11.2)	19 (11.2)	7 (5.6)	22 (28.1)	24 (30.9)	87
Total	76	76	38	190	209	589

During periods when the observer assessed crew task load to be close to saturation the pilots omit calling out 21 out of 40 (53%) of the occurring mode changes. In lower task load situations the shedding rate was 12% (see Figure 7).

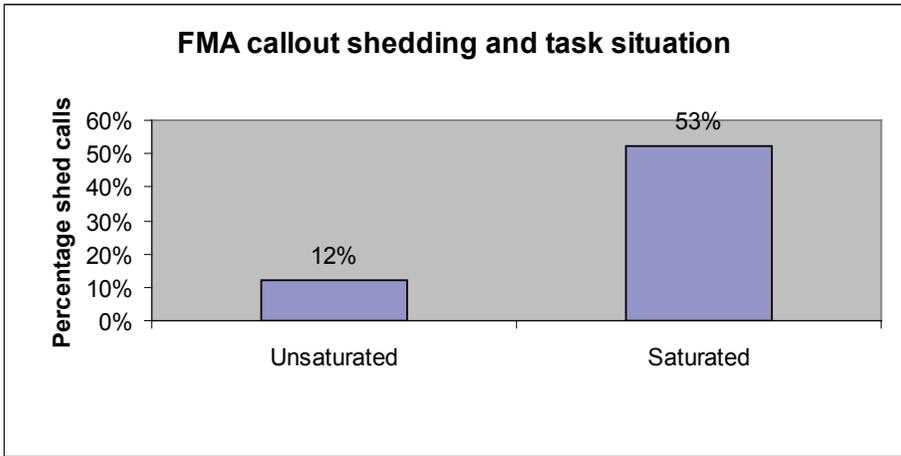


Figure 7: Percentage shed calls in a saturated versus unsaturated task load situation

In Table 5, we have cast the data in a contingency according to task load. Testing for significance with the χ^2 -test showed a significant difference between the groups, $\chi^2 (df = 1) = 44.64 p < .001$.

Table 5

Task Load

	<i>Unsaturated</i>	<i>Saturated</i>	<i>Combined</i>
Shed callouts	66 (81.1)	21 (5.9)	87
Called mode changes	483 (467.9)	19 (34.1)	502
Total	549	40	589

Discussion

More stringent callout procedures appear to correlate to less compliance, as evidenced by Operator A's 68% callout rate versus Operator C's 86% (against an average over all three operators of 85%). The up-front investment made by Operator C to think critically through which mode callouts are really important, and realistically can be expected to be followed, appears to generate a return in greater compliance. It is interesting to note that pilot experience with the A320 family aircraft seems to have no significant impact on the callout rate. While high task load does have a significant effect on callout shedding, it does not appear to be connected as clearly with phases of flight that are traditionally thought of as higher in task load (particularly descent and approach). In fact, descent and approach phases were associated with a lower callout shedding than takeoff and climb. The explanation of this apparent paradox may be an effect of the way task saturation was defined in this study. Task load in this study was estimated by a domain expert, well aware of normal task load during the various phases of flight. Higher task load situations may thus appear in other phases of flight than traditionally associated with high workload. Interestingly during high task load, other verbal coordination, especially that with Air Traffic Control, do not suffer. Mode callouts about the automation could be (and perhaps are) seen as a secondary task, while verbal coordination with, or about, other human partners in the system (e.g. the controller) are deemed central, or primary to the conduct of the flight.

There appears to be a floor effect with mode callouts: even at operator B (which required no mode callouts to be made on the A320 whatsoever) some mode changes are called out. These, interestingly, were callouts associated with vertical mode reversions that are connected to pilot changes in the lateral plan. The callouts thus made could be evidence of automation surprises, where the mode change is not directly attributable to a pilot action but rather a designed-in side effect that may come as unexpected to the crew (see Sarter & Woods, 1997). The floor effect is visible also at operator C (despite its critical up-front selection of which modes to call out); more modes get called out than what procedures specify.

Design assumptions of a flight manifested in the vertical flightplan with different phases of flight and switching conditions are regularly contested by the fact that during all observed flights of the A320 the pilots had to force the FMGS into the approach phase. In no flight did the pilots use the DES mode that lets the FMGS control the target descend speed and altitude constraints. When asked, the pilots stated that the uncertainty of the air traffic control intention while being radar vectored for final approach precluded the use of this high-level automation mode. They preferred to use the lower level automation OP DES mode, where the pilot controlled the descent constraints and the descent path more directly.

The normal flight in this study included both vectoring during the climbout, with subsequent reversions to the OP CLB mode. Also it contained several altitude level-offs during descent with speed restrictions and radar vectoring to position the aircraft in the approach sequence, again leading to reversionary mode changes from DES mode to OP DES mode. As discussed in the section above, this led the pilots to prefer to manage the flight in lower levels of automation than

the modes associated with the pre-planned (and most certainly not followed) vertical and lateral flightplan. It is interesting to note that the manufacturers assumed normal flight as it appeared in the flightcrew manuals, see Figure 8, contained an assumption of only 22 mode changes in a succession that was not observed in this study.

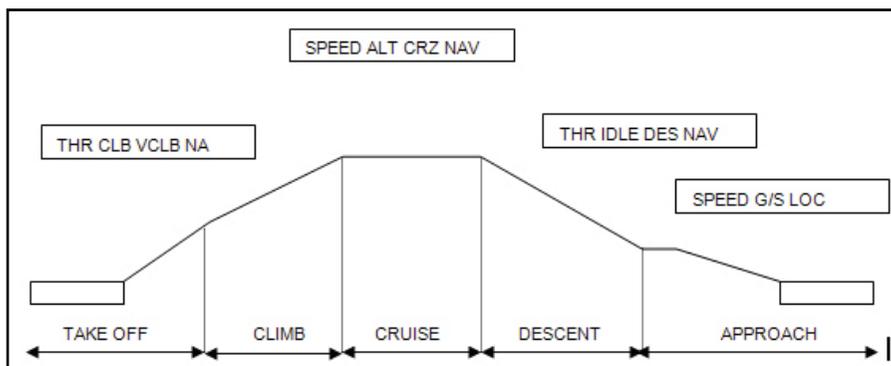


Figure 8. The assumed normal flight with typical flight guidance modes as shown in the Flight Crew Operating Manual

The study did not lead to conclusive claims about the relationship between callout shedding and flight crew performance. However, no automation surprises (see Sarter & Woods, 1997) were observed; whether mode changes were called out or not, did not seem to have a large effect on the potential for coordination breakdowns in the situations observed in this study.

Conclusions

The study reported here showed that on the Airbus A320 around 15% of required flight guidance automation mode change callouts are shed. There was strong influence of the number of required callouts on compliance. The more an airline company required its pilots to call out, the more callouts they will shed. Even if an airline does not require any mode callouts whatsoever, crews do call out some mode changes. These are the (unexpected) reversionary mode changes (such as open climb, or OP CLB) that were called out even if the operator did not require. This suggested a floor effect: some mode changes are so salient relative to crew expectations that they will provoke a callout independent of procedural imperative.

When modes were called out, pilots invariably used the correct verbiage; there was no improvisation. When callouts were shed, the crews studied here never revisited them (which, in contrast, they did do with briefings and checklists and other interrupted or uncompleted cockpit tasks). The study also showed a strong effect of task load on mode callout shedding, suggesting that mode callouts are seen as a secondary task relative to other aggregating priorities (e.g. verbal coordination with ATC is not shed in the same way, even during high task load situations). The study could not show any detectable effect of aircraft experience, suggesting that shedding of callouts is relatively stable across a pilot's

familiarity with the equipment, and that compliance with mode callouts hinges on factors other than experience.

This study could help operators become more sensitive to which mode callouts crews consider important given their operational context, and which mode changes crews can realistically be expected to call out. Cockpit procedure designers should note that FMA callouts are likely to be shed in exactly the situations where they were thought to be needed most. Cockpit procedures that capitalize on redundant sources to keep the pilots' aware of the state of the aircraft's automation such as briefings may thus add resilience to the autopilot-pilots triad. Training pilots to transition effortlessly between levels of automation is another possible tool for operators to ensure that pilots keep track of automation behavior. In the end, indeed, the FMA is probably an unsatisfying solution to ensuring crew awareness of automation status. Even such awareness of state (itself often incomplete and buggy, as has been pointed out previously [e.g. Sarter & Woods, 1997] and confirmed again here), does not necessarily enhance an understanding of automation behavior, which is, of course, the target for designers to get across and for crews to comprehend.

References

- Airbus Industrie (no year), *Getting to grips with approach-and-landing accidents reduction. 1.4 – standard calls*, Toulouse: Airbus Industrie
- Airbus Industrie. (2006). *Flight crew operating manual A320*, Toulouse: Airbus Industrie
- Anders, G. (2001). Pilot's attention during approach and landing: Eye- and headtracking research in an A330 full flight simulator. *Proceedings of 11th International Symposium in Aviation Psychology*. Columbus: Ohio State University.
- Björklund, C. M., Alfredsson, J. & Dekker, S. W., (2006). Mode monitoring and call-outs: an eye-tracking study of two-crew automated flight deck operations, *International Journal Of Aviation Psychology*, 16(3), 257–269
- Boeing Commercial Airplanes. (1999). *737 flight crew operating manual (Volume 1)*. Seattle: The Boeing Company
- Degani, A. and Wiener, E. L. (1994). *On the design of flight-deck procedures*. NASA Contractor Report 177642. NASA Ames Research Center.
- Dekker, S. W. A. (2003). Failure to adapt or adaptations that fail: Contrasting models on procedures and safety. *Applied Ergonomics*, 34, 233-238.
- Federal Aviation Administration (1996). *The interface between flightcrews and modern flight deck systems*. Washington, DC: FAA.
- Hüettig, G., Anders, G. & Tautz, A. (1999), Mode awareness in a modern glass cockpit – attention allocation to mode information, *Proceedings of the 10th International Symposium on Aviation Psychology*, Columbus, OH
- JAA (1997) *Joint aviation requirements, JAR-OPS 1*. Hoofddorp, the Netherlands.
- ICAO (2001). *Annex 2, Operation of aircraft*. Montreal, Quebec: International Civil Aviation Organization.
- Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2003). Concurrent task demands in the cockpit: Challenges and vulnerabilities in routine flight operations. *Proceedings of the 12th International Symposium on Aviation Psychology*. Dayton, OH.
- Mumaw, R. J., Sarter, N. B., & Wickens, C. D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. *Proceedings of 11th International Symposium in Aviation Psychology*. Columbus: Ohio State University.

- Nevile, M. (2004). *Beyond the black box: Talk-in-interaction in the airline cockpit*. Aldershot, UK: Ashgate
- Plat, M. & Amalberti, R. (2000). Experimental crew training to deal with automation surprises. In N. B. Sarter & R. Amalberti (Eds.), *Cognitive engineering in the aviation domain*. (pp. 280-307). Mahwah, Lawrence Erlbaum Associates.
- Sarter, N. B., & Woods, D. D. (1997). Teamplay with a powerful and independent agent: A corpus of operational experiences and automation surprises on the Airbus A320. *Human Factors*, 39, 553-569.
- Snook, S. A. (2002). *Friendly fire: The accidental shoot down of U.S. Black Hawks over Northern Iraq*. Princeton: Princeton University Press.

FINANCIAL STABILITY AND AIRLINE SAFETY: RELATIONSHIPS, CAUSES, AND CONSEQUENCES

Adam B. Schulman

1819 S Michigan Ave
Chicago, IL 60616
312-636-8886

a-schulman@kellogg.northwestern.edu

Abstract

This report explores the link between an airline's profitability and its safety record. Prior literature on the subject was reviewed and discussed to provide background on the subject and form a basis for the research. While there is an abundance of literature on this topic, conclusive evidence is still disputed among experts. The purpose of this study was to determine the relationship between profitability and safety using a new and distinct methodology from former studies. By thoroughly reviewing raw data from 1995 to 2004 for the top ten U.S.-based airlines, the author has conducted an independent analysis of the information and provided quantitative evidence that justifies the conclusions presented. A statistical analysis is included to validate the results. In addition, possible causation is discussed in detail, specifically investigating the extent that investments in safety projects and the level of maintenance outsourcing impact safety.

Introduction

This study considers the link between an airline's financial stability and its safety record. Past research indicates diverging views regarding whether this relationship indeed exists. Academic journals and contemporary periodicals were reviewed in-depth in order to assess, thoroughly, whether financially stable airlines have better safety records. This review suggested that a statistically significant correlation exists; therefore, this report explores the leading causes in-depth.

It is important to note that the existence of a correlation between two factors does not imply causation. Determining causation is difficult, as it is necessary to isolate certain factors and study them in a controlled environment. For the purpose of this study, two potential causes were considered: investment in safety enhancements and avoidance of low-cost, high-risk activities. Both of these are

managerial decisions that require a degree of balancing risk (safety) with fiscal stability. Arguably, other important factors should be considered, such as internal audit and quality assurance programs. It is clear that these programs have contributed positively to aviation safety. However, they were not included in this study due to the difficulty of statistically measuring the impact of these programs.

There are two fundamentally important aspects of the cost equation: making shrewd investments and effectively managing major expenses. Since airlines are able to exercise managerial judgment in both implementing optional safety enhancements and in selecting the degree of outsourcing, this study explored how these elements affect overall safety. The research presented strongly suggests that increased investment in safety enhancements yields a positive return in accident reduction. While it is expected that airlines adhere to mandatory Federal Aviation Regulations (FARs), implementing additional safety measures is up to the discretion of individual airlines. Some of these enhancements substantially reduce the risk of major accidents and incidents, but require initial financing. Many enhancements, such as the Traffic Collision and Avoidance System (TCAS) and the Terrain Awareness and Warning System (TAWS) were optional before the FAA made their usage mandatory. The enhancements currently under consideration have implementation cost, offset by additional risk reduction. This analysis is designed to examine how an airline's level of monetary investment affects safety. Investment is typically considered a current outlay of cash with an expected return in the future. Using this definition, this study assessed various safety enhancements that measurably improve airline safety and evaluated them in terms of both cost and risk reduction. Data and models from the Commercial Aviation Safety Team (CAST) have been applied to current and future safety enhancements to determine the cost impact to the industry. As mentioned, special attention has been given to the distinction between mandatory enhancements (required by the FAA) and optional enhancements (implemented at the airline's discretion).

The safety impact of increased maintenance outsourcing is a more contentious issue. Specific examples from both sides of the debate are presented to illustrate the major safety concerns that outsourcing may produce. The competitive environment has become so aggressive recently that drastic cost cutting measures are often necessary for an airline's survival. As mentioned in the Air Transport Association's (ATA) testimony before the Senate, spending on air travel has actually decreased in proportion to Gross Domestic Product (GDP). A snapshot of the industry reveals that airline revenue has diminished, while costs have increased (Air Transport Association, 2005). Most notably, fuel cost, the second largest expense after labor, is growing unpredictably. In some cases, the recent spike in fuel cost has outweighed that of labor. That fact limits the options of airline management to keep costs down. As a result, an industry trend towards lower-cost outsourcing is becoming popular. This report assesses whether such an action has an adverse effect on safety.

The New Environment

The business model for individual airlines has changed drastically over the past ten years. Low-cost airlines used to be known as "no frills" carriers, often providing limited service at prices that were significantly lower than the traditional

airlines. However, as the low-cost carriers (LCCs) began to grow profitably, major airlines needed to offer competitive fares to maintain their customer base. However, these airlines did not have the aggressive low cost structure of the LCCs, and therefore, could not profitably match these fares without reducing internal costs. In order to achieve these reductions, traditional airlines cut employee wages and benefits, while outsourcing a greater amount of maintenance to lower cost third parties. Concern over the safety implications of outsourcing has been a controversial issue, especially since the crash of ValuJet flight 592 in the Florida Everglades. Much has been written on both sides of this issue and more is expected in coming months, including a recent audit announcement from the Office of the Department of Transportation Inspector General (FAA, 2005). Some experts argue that the industry's excellent safety record proves that outsourcing does not present an additional risk, while others believe that this trend is a catastrophe waiting to happen (Fitzpatrick, 2004). This study has explored how the new ultra-competitive market has forced some carriers to change their operating plan and potentially take safety risks.

Some believe that this environment was created by deregulation in 1978. Deregulation provides a backdrop for examining the relationship between profitability and safety. During regulated times, airlines were given monopolies on routes throughout the United States. Air travel was limited to only the select segment of the general population that could afford it. The airlines were virtually guaranteed to earn handsome profits. When deregulation became law in 1978, there was a concern that the newly formed competitive market place would create an incentive for airlines to cut back on maintenance, safety initiatives, and regulatory compliance. Many airlines, such as Pan Am, Braniff, and Eastern, were unable to succeed in the deregulated environment. Others were forced to alter, drastically, their cost structure in order to compete effectively with the new low-fare entrants. Many studies have closely examined the effect that deregulation has had on airline safety. Historical data show that accident rates in general have been declining steadily since deregulation (Rose, 1992, p. 77). The risk of fatality was nearly three times less for the seven years after deregulation than for the seven years before it.

It is important to note the difficulty in isolating the factors that affect safety. New aircraft and technology improvements have contributed to the improved safety record since 1978. An attempt to draw a conclusion on whether the newly competitive market drove airlines to cut corners on maintenance would be nearly impossible. However, the data do not suggest that there is a drastic increase in accidents as a result of the new rivalries. In fact, 2002, 2003, and 2004 were three of the safest years for airline passengers ever recorded, yet those years presented some of the most difficult financial challenges to date (Borenstein, 2004). There are other influences affecting the data. Since September 11, 2001, the federal government has given cash infusions in the form of loans to struggling carriers. Though many reasons exist for this aid, including securing jobs, maintaining necessary air routes and stimulating competition, one must be sure not to overlook safety (McCartney, Carey, & Brannigan, 2001). An increased emphasis has been placed on airline security, with much of the cost being borne by the airlines. As a result, traditional safety improvement programs have taken a back-seat to the higher profile security measures. The expense of implementing new

safety technology and infrastructure is currently unaffordable to many of the top airlines. In fact, due to budget constraints and pressures on limited resources, the FAA evaluates safety initiatives in terms of return on investment (Pasztor, June 2003).

Limitations

This study is intended to provide evidence substantiating the link between an airline's financial and safety performance. Specific attention was given to determine causality; however, it is not intended to be all-inclusive. Many financial and non-financial factors contribute to the safety record of airlines. Two specific factors have been explored: investment in safety enhancements and level of maintenance outsourcing. Other causes have not been considered within the scope of this study.

The second limitation is that there was no restriction regarding the inclusion of the data in the incident database. If the FAA recorded an incident report during the study's time period, it was included herein without subjectively discriminating based on severity. As a result, minor incidents that may not be reflective of an airline's overall safety performance were included in the data set.

Review of the Literature

The Link between Financial Stability and Safety

There is renewed debate as to whether a link exists between an airline's financial well-being and its safety record. Intuitively, an airline that has greater financial resources will likely spend more on equipment, labor, and facilities. As mentioned earlier, there is evidence that supports both sides of this theory. It is a fact that airlines with limited capital cannot afford new airplanes, lucrative labor contracts or updated maintenance and training bases. However, what is unclear is whether this has a statistically significant impact on safety. In a free-market economy, an airline has many stakeholders with a vested interest in the company's operations. An airline, similar to other private enterprises, has an obligation to earn a profit and generate a return for its investors. As a result, airline management attempts to maximize revenue while minimizing cost. Though it is less visible to shareholders, safety is an essential part of this equation. If an airline suffers an incident or accident, it is very costly in terms of both dollars and reputation. There is no evidence to suggest that an airline's senior management would intentionally undercapitalize maintenance and safety programs. However, the research presented in this paper does suggest that many of the ways airlines cut costs involve reducing salaries of its workers and subcontracting a greater amount of maintenance work to third parties. An analysis of the impact of outsourcing is provided below in "Outsourcing – Benefits vs. Risks."

Other parties have a vested interest as well – specifically the FAA. It is the FAA's duty to regulate the airlines, ensuring that they are abiding by all regulations and operating their aircraft in a safe manner. Unfortunately, the FAA does not have the adequate resources to oversee the airlines properly. According to the DOT Inspector General's report,

While FAA has made progress in moving toward a more risk-based approach to safety oversight, FAA inspectors were not able to use the

oversight systems to monitor the rapidly occurring changes effectively. This is a significant concern in light of the fact that FAA is expected to lose about 300 aviation safety inspectors this year and in FY 2006 is only requesting budget authority to replace 97 inspectors. As a point of reference, there has been a lot of focus on hiring air traffic controllers—FAA has requested \$25 million to hire 1,249 new controllers during 2006, which includes 595 new positions. While that is a critical issue for the Agency, it is also important to maintain a safety inspector workforce that is sufficient to achieve its mission of safety oversight. Until its risk-based approach to safety oversight is operating effectively and targeting already constrained resources to the areas of greatest risk, FAA needs to determine if it can make enough efficiency gains in its operations to sustain the cut in staffing beyond 2005 (Department of Transportation Inspector General, 2005, p.i)

The general public has been made aware of certain cases where distressed airlines have been accused of taking shortcuts when it comes to safety. Since the safety of an airline is not readily transparent to the average person, the public relies on other factors to help determine whether an airline is indeed safe. There have been numerous studies undertaken to explore various relationships between safety and a more observable factor, such as financial health or service quality. If a strong correlation between a publicly recognizable factor and safety exists, then it can be effectively used as a proxy. Rhoades and Waguespack (1999) explored whether a connection exists between service quality and safety. The correlation was calculated, and it was determined that there is no significant correlation between service quality and safety. Their research therefore suggests that service quality cannot be used as a proxy for airline safety (Rhoades & Waguespack, 1999).

Since service quality has proven to be an unreliable indicator of safety, one can examine financial health for a possible relationship. Within the realm of finance, many approaches have been used to determine the financial strength of a company. Some models are based on credit rating, while others concentrate on liquidity constraints such as cash flow and working capital. This study has examined some of these approaches and provided the results of prior research as a basis for the analysis presented herein.

Previous research completed by Nancy L. Rose (1990) indicated that there is a strong link between profitability and product quality (safety). She used empirical evidence within the airline industry to support her claim. From 1957-1986, Rose found that operating margin was negatively correlated and statistically significant. Rose's formula predicted that a 7.6 % increase in operating margin resulted in a drop of the expected accident rate by 7.4 %. In cases of fatal accidents, the impact of operating margin was even more pronounced, though its statistical significance was less convincing. As Rose continued to analyze the sample further, she used airline size as a distinguishing factor, based on annual departures. Her results concluded that for small airlines, there is a strong and statistically significant relationship between profitability and safety. For medium-sized airlines, the effect was weaker and the data was less persuasive. More importantly, Rose found no clear relationship between operating margin and safety for large air-

lines. Therefore, based on Rose's research, the data as a whole supported her hypothesis; however, "efforts to divide the sample into separate carrier groups or separate time periods yield statistically inconclusive results" (Rose, 1990, p. 958).

Supporting Rose's conclusion on this lack of conclusive evidence is former Vice-Chairman of the National Transportation Safety Board (NTSB), Bob Francis. He claimed, "there is no evidence that there is a relationship between financial conditions and safety performance" (Fitzpatrick, 2004).

Conversely, a recently completed report by Noronha and Singal (2004) titled *Financial Health and Airline Safety* supported the claim that there is indeed a correlation. The authors used bond rating as a proxy for financial health and FAA data to represent safety. The rationale is that bond rating is an appropriate measure of financial condition because it is forward looking and relatively long term. In addition, bond ratings can be used to judge the likelihood of default, and consequently, the probability that an airline will have the incentive to decrease its investment in safety. However, as Noronha and Singal accurately pointed out, there are problems associated with using bond ratings – namely that these ratings are not updated as frequently as operating margins are reported. There is considerable debate as to which barometer is more accurate. Since bond ratings are typically higher for companies with sustained profitability, it is reasonable to substitute one for the other. Regarding the argument that bond ratings are forward-looking, it is unnecessary to glimpse into the future to determine if a link exists between profitability and safety. Since we cannot accurately predict safety mishaps, having a forward-looking instrument does not lend credibility to any derived correlations. Nevertheless, Noronha and Singal developed some important conclusions from their research. Principally, the authors discovered that a one-letter change in bond rating is consistent with a 10% change in the accident rate. Further, with respect to causality, they surmise that safety is dependent on financial health, rather than vice versa (Noronha & Singal, 2004).

Since there is ongoing discussion in existing academic discourse whether such a correlation exists, this study intends to research the data and form independent conclusions. The data studied in this report, which contain the financial and safety performance of the top ten U.S. airlines in terms of size (from 1995-2004), showed that a correlation does indeed exist.

Cost-Effective Safety Investments

While it may be debatable whether an airline's poor financial stability has an adverse effect on its safety record, it is widely accepted that prudent investments in safety programs lead to positive effects on its record. In the wake of the TWA flight 800 disaster, former Vice-President Al Gore was asked to lead a commission to study the future of aviation safety. On February 12, 1997, the report was delivered to President Clinton, recommending some specific targets and goals. Most notably, the Commission advocated that "government and industry should establish a national goal to reduce the aviation fatal accident rate by a factor of five within ten years and conduct safety research to support that goal" (White House Commission, 1997, p.1.1). As a result, the FAA announced a detailed plan to reduce the commercial aviation accident rate by 80% by 2007. The report fur-

ther requested, "The FAA should develop standards for continuous safety improvement, and should target its regulatory resources based on performance against those standards" (White House Commission, 1997, p. 1.2). The Commission recognized that the government needed to work together with industry in order to meet this ambitious goal. As a result, the Commercial Aviation Safety Team (CAST) was created. CAST is a cross-functional consortium of manufacturers, government agencies, airlines, and pilot unions. The organization is tasked with collecting data to provide recommendations for safety enhancements that will lead to the greatest reduction in the accident rate at the lowest cost. (CAST, 2004a).

Consistent with the goal of this study, CAST data has been used to determine the cost to industry to implement the safety enhancements that have the greatest impact on lowering the accident rate. In order to calculate the benefits and costs for a particular safety enhancement, a few factors need to be considered. The benefits consist of the injuries and deaths that are prevented as a result of the safety enhancement, in addition to the equipment losses avoided. This is a two-step calculation. First, one must determine the risk reduced as a result of implementing the safety enhancement. Then, in order to convert this benefit into monetary terms, one must acknowledge a cost for human life. The costs consist of the value of developing, purchasing and implementing the enhancement. When the benefits are divided by the costs (benefit/cost ratio), the outcome is an economic figure that can be used to compare various enhancements. The enhancements that return the greatest benefit (risk reduction) for the lowest cost (implementation) have been selected and prioritized. The data and results from CAST's research were examined in detail in order to provide the foundation for additional research into safety enhancements. As mentioned above, the main task of CAST is to reduce the accident rate 80% by 2007. In order to achieve this goal, the team studied the most common causes for accidents and selected six categories to group them: Controlled Flight Into Terrain (CFIT), Loss of Control (LOC), Uncontained Engine Failures, Runway Incursions, Approach, and Landing Weather (CAST, 2004a).

Data were then collected from previous fatal accidents (1987-2000) and classified into the above categories.

The CAST organization is divided into three groups: Joint Safety Analysis Teams (JSAT), Joint Safety Implementation Teams (JSIT), and Joint Implementation Measurement Data Analysis Team (JIMDAT). The JSAT is tasked with data analysis, while the JSIT uses this analysis to recommend safety enhancements. The JIMDAT is responsible for the master safety plan, including enhancement effectiveness, and future areas of study. (CAST, 2004a).

Using a relative Benefits/Costs Ratio (BCR) calculation, the benefits and costs of various enhancements have been compared and optimized. Given the tight budget constraints of the FAA and the airlines it regulates, it is more important than ever to get the greatest return on investment. Therefore, CAST sought enhancements that had the highest BCR. The benefits consist of the risk reduction multiplied by the cost of the lives saved, while the costs are the charges for implementing these enhancements. (CAST, 2004a).

The JSAT accumulated data on fatal accidents from various sources (including NASDAC, Airclaims, NTSB reports and FOQA data) from 1987 to 2000. Using this information, the accident data were categorized into one of the six categories mentioned above. Then, a causal analysis was performed, resulting in a collection of recommended enhancements that would have prevented the accident. These enhancements were turned into implementation strategies and were recorded into a database. Next, the enhancements were evaluated in terms of effectiveness and cost. Then the JSIT team prioritized the enhancements based on the greatest accident prevention at the lowest implementation cost. The result was a set of feasible enhancements that, if implemented, would reduce the accident rate substantially. The JIMDAT identified 46 of the most feasible enhancements and optimized them in order of risk reduction. The predicted result of these 46 enhancements is a 73% decrease in the commercial airplane accident rate. The 46 enhancements were selected based on the amount of risk reduced. According to the CAST information, of the 46 safety enhancements, 22 are complete and 24 are committed and in progress. Table 1, highlights the 46 enhancements selected (enhancements that were not selected for implementation were excluded) (CAST, 2004a).

Table 1a
CAST Enhancements

Status	Enhancement #	Safety Enhancement
Completed	1	CFIT TAWS – One Project
Completed	2	CFIT SOPs – One Project
Completed	3	CFIT PAI – Vertical Angles (PAI 1-7, 11)
Completed	4	CFIT PAI – VGSI at Runway Ends (PAI 8)
Completed	5	CFIT PAI – DME at Airports (PAI 2)
On plan	6	CFIT PAI – RNAV 3-D Instrument Approach (PAI 13-22)
On plan	7	CFIT PAI – RNAV RNP Instrument Approach Procedures (PAI 23-27)
On plan	8	CFIT PAI – XLS (ILS, MLS, GLS) (PAI 28-30)
Completed	9	CFIT MSAW – One Project
Completed	10	CFIT Proactive Safety Programs (FOQA + ASAP)
Completed	11	CFIT CRM Training
Completed	12	CFIT Prevention Training – One Project
Completed	13	CFIT ATC CFIT Training – One Project
Completed	14	ALAR Policies (Safety Culture) – CEO and DOS More Visible (1-2)
Completed	15	ALAR Policies (Safety Culture) – Safety Info in Manuals (3)
Completed	16	ALAR Policies (Safety Culture) – AFM Database for Inspectors (4)
Completed	17	ALAR Maintenance Procedures – Servicing Landing Struts (1)
Completed	18	ALAR Maintenance Procedures – Subcontractor Maintenance Guidance (2)
Completed	19	ALAR Maintenance Procedures – Policy on MELs (3) (Covers Recurring Maintenance Events)
Completed	20	ALAR Maintenance Procedures – DOS Internal Survey (4)
Completed	21	ALAR Flight Deck Equipment Upgrades – New Type Designs (1-3)
Completed	23	ALAR Flight Crew Training – One Project
On plan	24	ALAR Aircraft Design – Continuing Airworthiness (1-3)

Table 1b
CAST Enhancements (continued)

Status	Enhancement #	Safety Enhancement
On plan	25	ALAR Aircraft Design – Critical System Maintenance (4)
On plan	26	LOC Policies and Procedures – SOP – One Project
On plan	27	LOC Policies and Procedures – Risk Assessments and Management – One Project
Completed	28	LOC Policies and Procedures – Policies – Process to Inform Personnel/Flight Crew (1)
Completed	29	LOC Policies and Procedures – Policies – Flight Crew Proficiency Program (2)
On plan	30	LOC Training – Human Factors and Automation – One Project
On plan	31	LOC Training – Advanced Maneuvers – Implement Ground and Flight Training (1-3)
On plan	32	LOC Autoflight Design – New Designs (1-4)
On plan	34	LOC Displays and Alerting Systems – New designs (1-2) (Removed VSD)
On plan	39	LOC Basic Airplane Design – Icing (4-5) (scored zero for ground ice)
Completed	40	LOC Envelope Protection – New Airplanes (3)
On plan	46	RI Air Traffic Control Training – Enhanced Tower Controller Training (1-4)
On plan	47	RI – Tower Controller CRM Training (ATTE or similar)
On plan	49	RI SOPs for Ground Operations (1)
Completed	50	RI SOPs for Ground Operations for GA (1)
Completed	51	RI SOPs for Tow Tug Operators (3)
On plan	52	RI SOPs for Vehicle Operators (4)
On plan	53	RI Situational Awareness Technology for ATC – Enhanced Airport Surveillance Equipment (1-3)
Completed	55	RI ATC Procedures – SOPs for Controller Situational Awareness (1)
Completed	59	RI ATC Procedures – Readback requirement (5) (CAST decision is to implement the intent of this safety enhancement through policy rather than rule)
On plan	60	RI Pilot Training – One Project/SA, SOPs, CRM, All Resources
On plan	78	TURB Procedures for Reducing Cabin Injuries

The results of the CAST study showed that the greatest safety improvement is realized by implementing a Terrain Awareness and Warning System (TAWS) in all commercial airplanes. This device is critical to preventing CFIT accidents. CAST stated,

This safety enhancement substantially reduces or eliminates CFIT accidents by improving pilot situational awareness by establishing appropriate procedures for the installation and use of Terrain Awareness and Warning System (TAWS) equipment. Procedures include proper flight crew reaction in regard to TAWS aural and visual warnings (CAST, 2004a, p. 52).

As a result, the FAA issued a final rule in March 2000 requiring that all new commercial aircraft to be equipped with TAWS equipment by March 2003, and the entire commercial fleet to be equipped by March 2005. The regulation was added on March 29, 2001 (14 CFR §121.354).

The second part of the benefits equation is the cost of human life spared and equipment losses averted as a result of a safety enhancement. Clearly, this is not easily determined. In order to quantify something that is nearly impossible to measure, multiple sources were examined. The FAA recognizes \$3.0 million per fatality in the Economic Values for Evaluation of Federal Aviation Administration, Investment and Regulatory Decisions, June 1998, Report FAA-APO-98-8. That document also details the costs for serious and minor injuries. According to the FAA, a serious injury is equivalent to 1/17.4 (5.7%) of a fatality, while a minor injury is equivalent to 1/64 (1.6%) of a fatality. In addition, the FAA provides guidance of the costs of aircraft damage. A full hull loss is equal to \$17.4 million for a jet and \$3.84 million for a turboprop. Substantial damage averages 30% of a full hull loss and minor damage averages 5% of a full hull loss. The indirect impact of an accident is estimated to be four times (4x) the hull loss value, including lost business resulting from the accident. These values were used in the CAST's calculations for the economic value of an accident. Combined with the risk reduction estimate mentioned above, CAST can now financially predict the cost savings of safety enhancements (Federal Aviation Administration, 1998).

However, other organizations have placed different values on the cost of human lives. Some, such as the September 11 Commission, have offered settlements based on individual calculations of potential earnings and other factors. The U.S. Congress created the "September 11th Victim Compensation Fund of 2001," which was responsible for reaching settlements with the families of those fatally injured. Rather than paying a standardized sum, the administrator of the fund, Kenneth Feinberg, paid settlements based on "claimant's age, life expectancy income, marital status, and the number and ages of dependents." Feinberg said that "5,102 claims had been received as of that date, including 2,521 claims for decedents (85 percent of those eligible), and that the fund had disbursed nearly \$1.5 billion. Individual death compensation amounts have ranged from \$250,000 to \$6.9 million" (Romero, 2004). Another consideration is jury awards. One particular jury awarded over \$8 million to the estate of a 35-year-old marketing representative after he died in a plane crash. The victim was considered a top wage-earner, and the settlement reflected lost future earnings.

The wide disparity in settlements shows that this is not an exact science. However, the FAA number of \$3 million per fatality seems to be near the average of other calculation methods.

The final aspect to consider in the benefits/costs ratio is the economic impact of implementing these enhancements. This value includes the cost to the government, the manufacturers, and the operators. CAST studied each enhancement carefully and created a Statement of Work (SOW) to detail the implementation plan. Estimators then considered the cost of these enhancements, including, but not limited to, research & development to develop the technology, equipment cost for each airplane, airplane design and installation non-recurring, down time of

airplane to install, operating cost delta for airplane or airline, government and industry cost and training of flight crews (CAST, 2004a). All of these factors are considered key ingredients to the safety enhancement selection process.

Outsourcing – Benefits vs. Risks

As mentioned above, there is considerable debate as to whether an increase in maintenance outsourcing leads to a decrease in airline safety. Research does not provide conclusive evidence one way or the other. However, a review of industry journals and contemporary periodicals provides a wealth of information on both sides of the debate. On one side, the FAA and the Air Transport Association (ATA) – a trade group representing major airlines – argue that the industry’s excellent safety record proves that an increase in outsourcing does not translate into an increase in safety incidents. The ATA has publicly stated, “There is no distinction, statistically, when you talk about safety, between work performed within the airline and work performed by a third party maintenance provider” (Griffin, 2005). When asked about this impact Francis, former vice chairperson of the NTSB, said, “There is no evidence that there is a relationship between financial conditions and safety performance” (Fitzpatrick, 2004). However, some of Francis’ former colleagues disagree. Former NTSB member, Goglia, is concerned with the degree of airline outsourcing, stating, “It’s particularly worrisome as the percentage of outsourced maintenance climbs” (Mecklar, 2005). Furthermore, in an interview with CNN, former NTSB chairperson Hall cited the crash of a USAirways Express commuter plane that crashed due to substandard work done by an outsourced maintenance company (Griffin, 2005). This accident investigation revealed gaps in the maintenance oversight of a third party company. The accident report showed that the mechanics, inspectors, and safety supervisors all worked for different companies, while USAirways Express provided little hands-on supervision of their subcontractor. While there were other factors in the crash, such as improperly loading the aircraft beyond gross takeoff weight, it is important not to overlook the chain of events that led to the crash. A deeper examination showed that there was no accountability for the subcontractor’s work, leading to blurred responsibilities and improper training of its employees. It was later discovered that the mechanic that improperly adjusted the cable had never done this procedure before and had missed some of the steps involved (NTSB, 2004).

Within the FAA, continued debate between administrators and inspectors is ongoing. A recent Department of Transportation Inspector General (DOTIG) report raised some significant questions about the ability of the FAA to effectively regulate and monitor maintenance contractors. Some issues cited in the report were the increase in unmonitored work performed at night, the lack of attention given to the safety of new procedures driven by airline economics and the shortage of FAA inspectors required to perform the necessary inspections. Budget cuts are forcing the FAA to reduce its staff of inspectors. In 2005, 300 inspectors were expected to retire, and only 97 will be replaced (Department of Transportation Inspector General, 2005). This is especially alarming considering that many planned inspections are not completed due to a lack of resources.

Another concern of safety experts is not necessarily the percentage of outsourced maintenance, but rather the specific companies that are selected for maintenance. It has always been an accepted practice to outsource maintenance

to the original manufacturer of certain parts, such as engines. Few would argue that this poses a safety threat – in fact, it is likely safer to send these parts back to their manufacturer, where mechanics and engineers are expertly familiar with the maintenance procedures (Carey & Frangos, 2005). It is less clear whether outsourcing to non-manufacturer third party shops is as prudent. Research suggested that many of these companies hire non-certified mechanics to perform much of the maintenance. These mechanics are less expensive to employ and therefore result in a less costly operation than traditional maintenance programs at major airlines. Fewer than one in ten workers are FAA-certificated Aircraft Mechanics (Finnegan, 2005). Typically, these mechanics are there for oversight and guidance, rather than actually performing the work. In some cases, maintenance work is done outside of the United States. For example, JetBlue Airways will have 17 of its 68 Airbus A320 aircraft flown to El Salvador for heavy maintenance. In an attempt to save money, JetBlue has non FAA-certificated mechanics perform much of the work for approximately half of what it would cost in the United States (Carey & Frangos, 2005). These mechanics are licensed by the Salvadorian Aviation Authority. JetBlue is not alone. United Airlines is doing maintenance work in China, and Northwest is using companies in Singapore and Hong Kong. American Airlines, on the other hand, outsources very little work to third parties. They prefer to keep maintenance in-house where it has “much more control of the whole repair process” (Carey & Frangos, 2005).

It may be statistically difficult to prove that this type of outsourcing leads to a decrease in safety. Some evidence points to a looming disaster, while other studies discredit the threat. A team of researchers from Embry-Riddle Aeronautical University performed a study on the correlation between maintenance spending and maintenance quality (safety) (Rhoades, Reynolds, Waguespack, & Williams, 2005). They concluded that increased spending on maintenance is associated with increased line maintenance activity and therefore increased overall safety. Using Service Difficulty Reports (SDRs), the researchers were able to use them as a proxy for safety related maintenance activity. They then compared the SDRs to the maintenance spending and calculated a correlation. While they concluded that the level of spending and the accident rate are negatively correlated, they added a caveat to their results. Factors such as the number of aircraft types and average age affect spending, but not necessarily safety. For example, Southwest Airlines flies only one type of aircraft, resulting in low maintenance expenditures. Nonetheless, Southwest has an excellent safety record. The authors also pointed out that their study was limited in nature, especially since some FAA reporting is subject to interpretation. They suggested an increase in the number of inspections of outsourced and foreign repair stations to ensure standards are being met.

Research on High Reliability Organizations (HROs) performed by Roberts may lend insight into airline maintenance activities. According to Roberts, HROs tend to exhibit additional redundancy, which leads to safer operations. By training multiple people on a particular task and training a single person on multiple tasks, the organization has more built-in redundancy. For example, a third-party maintenance shop may specialize in a certain type of work while the airline’s in-house facility may not. In this case, it may be more advantageous to outsource the work (Roberts, Bea, & Bartles, 2001).

Methodology

The Link between Financial Stability and Safety

This study considered previous research completed on the issue of whether an airline's safety is impacted by its financial stability. The established data presented in the "Literature Review" above proved to be inconclusive in reaching a consensus among safety experts. Therefore, using a new set of raw data, the author completed extensive research that has provided an additional perspective.

The first part of the study was to gather information on airline incidents, financial reporting, and fleet usage. The nature of this research demanded comprehensive data on all three of these topics. While the purpose of obtaining incident and financial data is clear, the importance of fleet usage cannot be overlooked.

The financial information is based on publicly available resources, namely accounting income statements. Using Mergent Online (<http://www.mergentonline.com>), the income statements for the top ten US based airlines¹ were reviewed for the ten-year period from 1995 to 2004. A listing of these airlines with their respective total flight hours is provided below in Table 2.

Table 2
Top Ten U.S. Airlines Based on Flight Hours

1995 - 2004	
Airline	Total Hours
American	21,639,595
United	18,141,757
Delta	16,910,307
Northwest	11,620,733
Continental	10,495,821
Southwest	10,447,442
US Airways	10,123,185
America West	4,087,622
Alaska	3,104,994
ATA	1,527,229

The profit margin used is calculated as total net income divided by the total operating revenues. The net income calculated is after tax and after any extraordinary expenses. The total operating revenue includes passenger, cargo, and mail revenues. Profit margin has been used as the main indicator for financial stability. Past studies used bond ratings as indicators for financial stability, but as mentioned previously, there are problems associated with using bond ratings. Profit margins are updated more frequently and are closely correlated with bond ratings, making it a better choice for a barometer of financial health.

¹AirTran was excluded from this study because the airline did not include the operations of ValuJet in 1995 – the first year of the sample.

The safety data are derived from The National Aviation Safety Data Analysis Center (NASDAC) - *FAA Accident /Incident Data System (AIDS)*. A search was performed of all accidents and incidents from January 1, 1995 through December 31, 2004 for the ten airlines mentioned above. There was no subjectivity added to this data set. The incident numbers reflect purely the numbers from the NASDAC system. No distinctions were made for the severity of the incident. An incident of any magnitude is a safety infraction, and this study focuses on total safety breakdowns, not just fatal accidents. In addition, incidents were not examined for cause or fault. Therefore, it is important to note that there are incidents included in this data set that may not be typically considered an airline's responsibility. To prevent unintentional bias, these results were not excluded from the study. The author has not detected a pattern among the non-safety related incidents that affect one airline more than another does. It appears that these incidents affect all airlines equally and do not measurably influence the results.

Fleet usage data were obtained from the Airclaims CASE database. This system tracks fleet usage for all airlines and is updated monthly. For each of the ten airlines examined, CASE was used to collect data on fleet hours and cycle time. The importance of this information rests in its ability to normalize the data set. It would be unfair to punish a larger airline for having more incidents simply because its operations are more sizeable. After considering both flight hours and cycles, the author chose to use cycles as it serves as a better normalizing tool. Since it is widely known that accidents are more likely during the takeoff and landing phases of flight, cycle data are a more accurate barometer than flight hours. Cycle time was gathered on the ten airlines' fleets from January 1, 1995 through December 31, 2004.

Using the raw data on profitability, incidents, and fleet usage, a score was calculated to reflect the number of incidents per cycle during this test period. In all cases, this number was infinitely small. To ease interpretation and comparison of the data, the calculated score was multiplied by a factor of 1,000,000.

$$score = \left(\frac{\#totalincidents}{\#totalcycles} \right) \times 1,000,000$$

This calculated score was then compared to the average profitability of the airline over the ten-year period. The correlation between these two factors is discussed in the "Results" section. In order to use these results as a predictive tool, statistical data need to be calculated. A standard exponential regression is the best-fit line. The exponential regression has more predictive power than the linear regression model. This trend line, along with the correlation coefficient and R2 statistic was calculated and included in Table 1.

The Effect of Safety Investments and Outsourcing

The research performed in the section above only studies correlation, not causation. In order to explore potential causes, one must review safety investments and outsourcing in order to examine their influence on safety. Ideally, data would be collected on an individual airline's investment in safety enhancements and compared to their safety performance. Unfortunately, this information is not publicly available. Balance sheets typically show short-term and long-term invest-

ment, however, there is no way to segregate the data further. Since these categories include investments in areas other than safety, they cannot be used to derive any conclusions about the impact of such investments. Therefore, one must rely on the research performed by the CAST to prove that additional investment in safety enhancements leads to an improvement of an airline's safety record.

As mentioned in the earlier sections, the effect on outsourcing is far less clear. Since no prior research was found that specifically examines the correlation between level of outsourcing and safety, the author used raw data to derive a conclusion. The Inspector General of the Department of Transportation has accumulated the percentage of maintenance work outsourced by the top U.S. airlines in 2002 and 2004. Using an average of these percentages for each airline yields the average level of outsourcing over the three-year period 2002-2004 (Meckler, 2005). By plotting this number against their incident/cycle score calculated over the same time period, a trend line and correlation coefficient can be calculated. These statistics can be used to measure the effect that outsourcing has on safety. If there is a correlation, it may indicate that outsourcing is a contributor to the overall correlation between profitability and safety.

Results

The goal of this project was to explore whether a correlation exists between an airline's financial stability and its safety record. Using the raw data explained earlier, the author is able to draw conclusions on the strength of this relationship and suggest possible causes.

The data in Table 3 show a correlation between profitability and incident score of -0.6. Correlation is measured on a scale of -1 to +1, with a correlation of -0.6 considered relatively strong. This correlation is statistically significant with a P-value of 0.00002.

Table 3
Incidents per Cycle

1995 – 2004					
Airline	Average Profitability	# of Incidents	Total Cycles	Inc/Cycles	Incident Score
Alaska	1.0%	45	1,697,270	0.0000265	26.5
American	-1.2%	243	8,596,824	0.0000283	28.3
American West	-0.8%	51	2,083,163	0.0000245	24.5
ATA	-6.5%	64	580,908	0.0001102	110.2
Continental	1.8%	202	4,382,395	0.0000461	46.1
Delta	-2.9%	360	8,792,925	0.0000409	40.9
Northwest	-0.6%	220	5,774,816	0.0000381	38.1
Southwest	7.8%	130	8,639,359	0.0000150	15.0
United	-4.4%	225	7,415,194	0.0000303	30.3
USAirways	-3.9%	184	6,525,725	0.0000282	28.2
Correlation: -0.6					

The regression formula that represents the best-fit line is $Y = 30.925e^{-8.8432x}$. This formula yields a trend line that has an R-squared of 46%. This means that 46% of the variance in X (profitability) can be explained by variation in Y (incident score). Likewise, 46% of the variance in Y (incident score) can be explained by (or goes along with) variation in X (profitability). More simply, 46% of the variance is shared between both of these factors. The data have been plotted and included below in Figure 1.

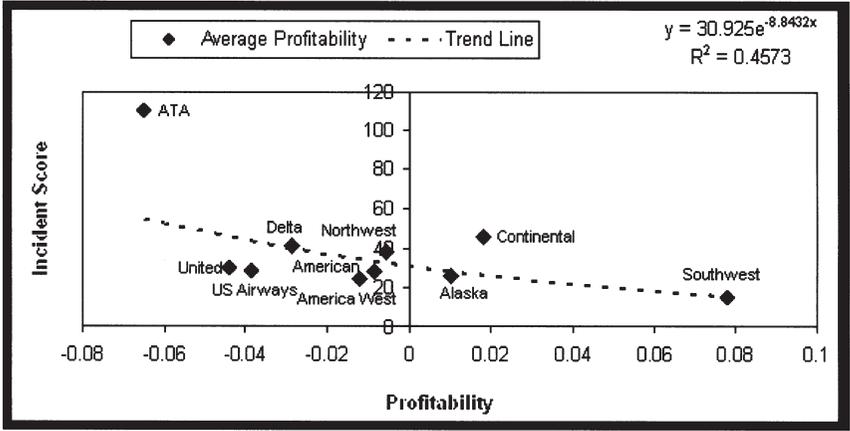


Figure 1. Profitability vs. Incident Score (1995-2004).

The data suggested that financial stability has a direct impact on safety, though it is not the only factor that affects it. There are a number of factors that can influence the results, such as route networks, types of aircraft, and fleet utilization. The data have not been adjusted to account for possible variations due to these factors. Nonetheless, it appeared that financially stable airlines tend to have good safety records. However, by studying the variation from the regression line, it was clear that some airlines have safety records better than what the model would suggest, while some have worse. Most notably, United Airlines and U.S. Airways have better safety records than their financial condition would suggest (both carriers are in bankruptcy protection at the time of this writing). Conversely, Continental Airlines has a safety record worse than what its profitability level would suggest. A potential anomaly in the data set is ATA. However, after further examination, this data point is an extreme value rather than an outlier. Whereas outliers should be removed from the data set, extreme values tend to improve the accuracy of the model. ATA is nearly off the charts when it comes to profitability (or lack thereof) and safety. In both categories, this carrier far exceeds the results collected from other companies.

Now that the relationship between profitability and safety has been established, the factors that cause it must be examined. Since investment in safety enhancements and level of outsourcing are two areas under airline management control, they have been studied as two potential contributors to this relationship.

CAST has extensively studied safety enhancements for the entire industry and has made specific recommendations to the FAA to implement those enhance-

ments that have the greatest return on investment. However, there are additional enhancements that have not yet been implemented and are currently at the discretion of the airline. Using data developed by CAST, one can calculate the return on investment for a particular enhancement. For example, an airline may examine the business case for installing a turbulence detection system. Any encounter with turbulence that results in injuries is counted by the FAA as an incident. Therefore, if an airline wants to reduce its number of incidents, focusing on avoiding turbulence would yield a positive improvement to its safety record. However, the current financial condition of many U.S. airlines prevents investment in costly safety projects. A cost/benefit analysis can be performed to determine whether this is a worthwhile investment. The adverse effects of turbulence can be measured by calculating the cost of related fatalities and injuries. The CAST reports that a fatality costs \$3,000,000, a serious injury is \$172,414, and a minor injury costs \$46,875. The Department of Transportation provides an average of \$192,000 per injury (OST, 2002). Given these cost projections, the total cost of injuries during the study period (1987-2000) is equal to approximately \$62 million. Therefore, the average injury cost is about \$4.8 million per year. If unreported injuries are included, this projection could increase up to an additional \$22 million (OST, 2002).

The cost of implementing a graphical display on new and in-service aircraft has been estimated by JSIT. CAST has identified four enhancements (Table 4 - numbered 71-74) of which the total cost is estimated to be approximately \$49 million. However, approximately 1/4 of these expenditures are expected to be borne by the U.S. government. That leaves manufacturers and airlines left with an investment of \$38 million. Depending on whether the unreported injuries are considered, the payback period varies. Assuming these are not included, the investment is recovered in approximately 8 years. This is considered a conservative estimate, as it does not include the cost for aircraft damage, maintenance inspections, and delays. If these costs are factored in to the equation, the payback period is significantly shorter.

Table 4.
Turbulence Enhancements.

Status	Enhancement	Safety Enhancement
Not currently on plan	71	71. TURD Graphical displays - Carry On
Not currently on plan	72	72. TURD Graphical displays - Panel Mounted - New Production
Not currently on plan	73	73. TURD Airborne Detection - Enhanced Radar-New Production
Not currently on plan	74	73. TURD Airborne Detection - Enhanced Radar-Retrofit Windshear-Equipped Aircraft

Using turbulence as an example, it becomes clear that an individual airline's investment in safety enhancements leads to a better safety record as well as a long-term financial benefit. Some other initiatives that are not mandatory but suggest better safety records are heads-up displays (HUDs), moving map displays and digitally enhanced data links for ATC instructions (CAST, 2004b).

A second potential cause for the correlation between financial stability and safety is the level of outsourcing. As mentioned above, there is little agreement among experts on whether the level of outsourcing affects safety. As shown in Figure 2, the trend line between safety and level of outsourcing is negatively related.

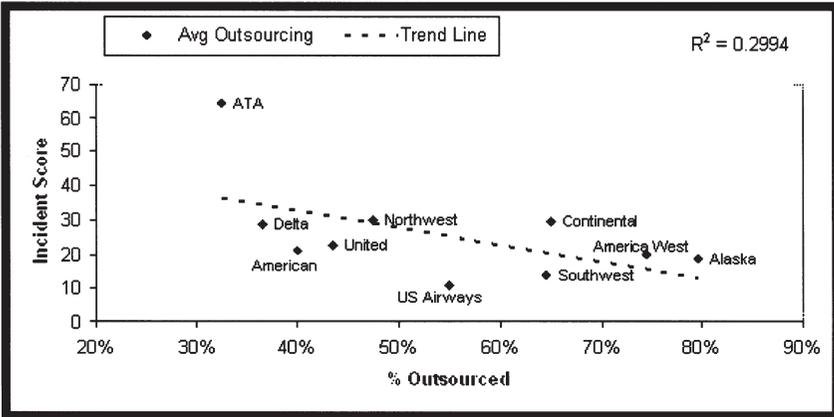


Figure 2. Level of Outsourcing vs. Safety (2002-2004).

There is a trend nowadays to send maintenance work out of the country, to places like El Salvador, Hong Kong, and China. While this particular study does not attempt to distinguish between in-country and foreign repair work, it does examine whether outsourcing in total affects safety. Table 5 shows a correlation between outsourcing and safety of -0.5 and R-squared of 30%. Contrary to conventional wisdom, the data indicate that as the level of outsourcing increases, the incident score decreases. There are a few important factors to note. Similar to the data in Table 3, ATA appears to be an extreme value. Its average outsourcing rate is the lowest, while its incident score is the highest. This may be skewing the data considerably. Secondly, the low R-squared indicates that only 30% of the variance in incident score is explained by outsourcing. Alternatively, if the ATA data point is treated as an outlier, the R-squared decreases to 12% and the correlation becomes -0.4.

A possible reason that higher outsourcing results in better safety records is related to profitability. As the data provided has suggested, profitability and safety are correlated. In addition, as provided in Table 6, profitability and outsourcing are also correlated. The powerful positive correlation between profitability and outsourcing (+0.8) suggests that the most profitable airlines also outsource the greatest amount of work to outside subcontractors. Given that airlines can save up to 70% on maintenance labor by outsourcing work, it is clear that this will improve their bottom line. If, as the data suggests, this increased level of outsourcing does not negatively affect safety, the airline is able to spend the captured savings on other aspects of its operation. Consequently, the airline may be able to purchase newer aircraft, improve pilot training, or invest in other safety enhancements.

Table 5
Outsourcing and Safety

Airline	Incidents ('02-'04)	Cycles ('02 - '04)	Score	% of Out-sourcing	% of Out-sourcing	Avg Out-sourcing	Profit-ability
Alaska	10	536,032	18.7	79%	80%	80%	-1.8%
American	57	2,710,181	21.0	38%	42%	40%	-10.5%
American West	12	601,252	20.0	77%	72%	75%	-6.7%
ATA	15	233,052	64.4	22%	43%	33%	-16.6%
Continental	34	1,146,823	29.6	65%	65%	65%	-2.9%
Delta	51	2,133,828	28.6	38%	35%	37%	-16.7%
Northwest	39	1,707,169	29.9	44%	51%	48%	-6.3%
Southwest	39	2,859,252	13.6	65%	64%	65%	5.5%
United	41	1,806,699	22.7	33%	54%	44%	-17.8%
USAirways	15	1,364,473	11.0	50%	60%	55%	-11.8%
Correlation:	-0.5						

Table 6
Outsourcing and Profitability

Airline	Avg Out-sourcing	Profit-ability
Alaska	80%	-1.8%
American	40%	-10.5%
American West	75%	-6.7%
ATA	33%	-16.6%
Continental	65%	-2.9%
Delta	37%	-16.7%
Northwest	48%	-6.3%
Southwest	65%	5.5%
United	44%	-17.8%
USAirways	55%	-11.8%
Correlation:	0.8	

Conclusion

Since the American airline industry has become so competitive, traditional airlines no longer have the market power to sustain high fares. Today's environment favors low-cost carriers because they can set fares at a minimal level and still turn a profit. In order to compete, traditional airlines have been required to match these fares and cut costs in order to remain solvent. While all major airlines used to do their own maintenance, low-cost carriers took advantage of the less expensive third-party subcontractors. This new environment has caused all airlines to reexamine their business plans to squeeze cost out where possible. It is therefore critically important to ensure that no additional risk to the traveling public is created as a result of operating in this low cost environment.

This project explored the link between financial stability and safety. The research presented indicated that there is indeed a positive correlation between

these factors. Further examination suggested that this relationship could be explained by a profitable airline's ability to invest in safety enhancements, invest in new equipment, and attract and retain talented employees. There may be other reasons as well, perhaps some that are intangible. Wages and morale may be higher at profitable companies, thus driving employees to perform better at their jobs. Conversely, airlines in financial trouble may resort to labor reductions, creating a larger workload and more stress for the remaining employees. Less profitable airlines may be unable to provide the same wages and benefits as the more solvent ones, resulting in high turnover and unhappy employees.

While the reasons mentioned above serve to explain the correlation between financial stability and safety, one important factor does not contribute. It appeared that a high percentage of maintenance outsourcing does not affect an airline's safety record negatively. In fact, based on the research presented, the more an airline outsources, the better its safety record. As previously mentioned, an explanation for this could be that both factors are highly correlated with profitability.

There has been a considerable amount of research done on this topic, yet the existing literature has not convinced some experts of the veracity of this claim. While the data presented made a strong case for believing that this relationship exists, more research needs to be done into the underlying causes. Though speculation into these causes may continue for the foreseeable future, one thing is clear: promoting a culture of safety must remain a top priority for airlines worldwide, regardless of their financial situation.

Future Studies

This project examined some of the economic factors that contribute to airline safety; however, it provided only a limited causal analysis. Future research can build on the concepts presented herein and explore additional factors that affect overall safety. There are numerous aspects of an airline's financial operations that can be investigated for a causal relationship, such as fleet age and crew training. Do airlines with greater financial resources buy more modern aircraft with advanced safety features? Do fiscally sound airlines spend more on their training budget? These questions were not addressed within this study and deserve further attention.

References

- Airclaims CASE Database (version 2.0.3) [Computer Software.] London, England.
- Air Transport Association. (2005, July 13). Hearing on airline financial stability. Statement for the record before the *Committee on Commerce, Science & Transportation Subcommittee on Aviation United States Senate*: James May.
- Alexander, A. & Mecia, T. (2005, June 9) FAA safety inspections for troubled, low-cost airlines deficient, report says. *Knight Ridder Tribune Business News*, p. 1.
- Alexander, A. & Reed, T. (2004, January 8). Charlotte, N.C., plane crash prompts changes in national maintenance oversight. *Knight Ridder Tribune Business News*, p. 1.
- Borenstein, S. (2004, January 3). 2003 ranks as the safest for airline industry. *The Seattle Times*, p.1

- Carey, S. & Frangos, A. (2005, January 21) Airlines, facing cost pressure, outsource crucial safety tasks. *The Wall Street Journal*, p.A1
- Commercial Aviation Safety Team (CAST) (2004A). CAST process overview. *NTSB Air Cargo Safety Forum*.
- Commercial Aviation Safety Team (CAST) (2004B). Plan and metrics. *NTSB Air Cargo Safety Forum*.
- Department of Transportation Inspector General (2005). *Safety oversight of an air carrier industry in transition*. (Report Number: AV-2005-062).
- Federal Aviation Administration (1998). *Economic Analysis of Decisions -- Revised Guide*. (Report Number FAA-APO-98-4).
- Federal Aviation Administration (2005). *Audit announcement—air carriers' outsourcing of aircraft maintenance*. (Project No. 05A3001A000): David Dobbs.
- Finnegan, B. (2005, March). Outsourcing Airworthiness: A question of certification and span of control. *Aircraft Maintenance Technology*. 44-45
- Fitzpatrick, D. (2004, November 7). Evidence suggests airlines' financial woes don't impact safety. *Knight Ridder Tribune Business News*, pg. 1
- Griffin, D. (2005, March 5). Rise in outside repairs raises questions. *CNN*. Retrieved October 19, 2006 from <http://www.cnn.com/2005/TRAVEL/03/03/airline.maintenance/>
- Lin, C. (2005, February 20). *Secret Tapes of George W. Bush; Is Airline Outsourcing a Safety Threat?* [Television Broadcast] CNN.
- McCartney, S., Carey, S. & Brannigan, M. (2001, November 6). Capital flight: as big losses widen, an airline shake-up appears unavoidable. *The Wall Street Journal*, A-1.
- Meckler, L. (2005, June 9). Airline turmoil poses safety issues; government report says financial troubles and outsourcing of maintenance create risks. *The Wall Street Journal*, p. D1.
- National Transportation Safety Board. (2004, February 26). *Aircraft Accident Report - Loss of Pitch Control During Takeoff - Air Midwest Flight 5481*. Retrieved October 19, 2006 from <http://www.nts.gov/publicctn/2004/AAR0401.htm>
- Noronha, G. & Singal, V. (2004). Financial health and airline safety. *Managerial and Decision Economics*, 25(1),1-16.
- OST (Office of the Secretary of Transportation) (2002). *Memorandum, "revised departmental guidance: treatment of value of life and injuries in preparing economic evaluations*: Carl Swerdloff.
- Pasztor, A. (2003, May 21). Crash probe reveals maintenance gaps. *The Wall Street Journal*, p. D3.
- Pasztor, A. (2003, June 18). Crash avoidance takes back seat to terror fears – airlines sideline projects geared at preveneting common accidents as spending on security rises. *The Wall Street Journal*, D-1.
- Rhoades, D. & Waguespack, B.(1999). Better safe than service? The relationship between safety and service in the us airline industry. *Managing Service Quality* 9(6), 396-401.
- Rhoades, D., Reynolds, R., Waguespack, B., & Williams, M. (2005). The effect of line maintenance activity on airline safety quality. *Journal of Air Transportation*, 10(1), 58.
- Roberts, K., Bea, R., & Bartles, D. (2001). Must accidents happen? Lessons from high-reliability organizations. *The Academy of Management Executives* 15(3), 70-78.
- Romero, J. (2004). A victim's eye view of the september 11th victim compensation fund. *Defense Counsel Journal*, 71(1), 64-70.
- Rose, N.L. (1990). Profitability and product quality: Economic determinants of airline safety performance. *The Journal of Political Economy* 98, 944-964.

Rose, N. L. (1992). Fear of flying? Economic analyses of airline safety. *The Journal of Economic Perspectives* 6(2), 75-94.

White House Commission on Aviation Safety and Security (1997). *Final report to President Clinton*: Gore, A.

A Case-based Review of Critical Incidents in General Aviation for Improved Safety

Jason J. Saleem

*Center for Engineering Learning and Teaching
University of Washington
Seattle, WA, 98195-2183, USA
Phone: (206) 919-1184
E-mail: jason.saleem@gmail.com*

and

Brian M. Kleiner

*Grado Department of Industrial and Systems Engineering
Virginia Polytechnic Institute and State University
Blacksburg, VA, 24061-0118, USA
Phone: (540) 231-4926
E-mail: bkleiner@vt.edu*

Abstract

We report on critical incidents in which pilot error occurred during field observations of landing approaches to a mid-sized, controlled airport. These occurrences included a case where a hand-off of air traffic control (ATC) from the airport approach to airport tower was delayed by the pilot, a case where the pilot requested an incorrect runway for a practice instrument approach, and a case where a pilot missed a course correction command from ATC. Each of these cases involves communications errors, where information exchanged from ATC to the pilot was not initially processed by the pilot or the information exchanged was incorrect. We use these cases to demonstrate how modest, pilot-focused design enhancements to the GA human-machine system may help minimize the occurrence of similar errors.

Introduction

The field of human factors has a rich tradition in aviation, and continues to help inform the design of aviation technologies to meet the demand of future aviation systems. Aviation systems should help minimize 'pilots errors,' while also helping pilots diagnose and recover from errors. This paper presents cases of pilot error in General Aviation (GA) recorded in the field and shows how human-centered, aviation technologies may help minimize the occurrence of similar

errors. This study was part of a larger research endeavor that examined human-machine performance in a GA environment using a combination of laboratory experimentation and field study (Saleem, 2003; Saleem & Kleiner, 2005). Three cases of pilot error in this paper were unanticipated events that occurred during the field observations from this larger research effort. The potential for disaster was especially high for the first two cases and reporting on these incidents may help prevent similar incidents from occurring. One case involved a potential runway incursion, which is particularly relevant considering the recent case in Lexington, KY, where an aircraft used the incorrect runway for take off and crashed shortly after (National Transportation Safety Board, 2006). The purpose of this paper is to present these three unexpected incidents that occurred during data collection, describe them in a way that is consistent with the literature on aviation error, and discuss potential ways they may be addressed in system design.

Field observations can give a realistic view of the full complexity of a work system, uncover the cognitive and collaborative demands imposed by a domain, and guide the development of new types of support systems (Roth & Patterson, 2005). Presenting and analyzing the pilot errors that occurred in this study may inform how modest design enhancements may assist pilots in similar aviation scenarios. This type of research is especially relevant today, as GA moves toward a concept of 'free flight', and pilots will face unanticipated challenges and will need to be supported with resilient systems (e.g., Hollnagel, Woods, & Leveson, 2006) to help minimize error, and facilitate recovery from errors when they occur.

Methods

We employed naturalistic, or ethnographic observation (e.g., Roth & Patterson, 2005; Stanney & Maxey, 1997) of the pilot performing an instrument flight rules (IFR) and visual flight rules (VFR) landing approach with a Cessna 172 to the Roanoke airport, a controlled airport with an instrument landing system (ILS). The Roanoke airport was also used as the airport in a flight simulator study which reported on pilot performance, workload, and situation awareness (Saleem & Kleiner, 2005) and was the focus of a joint field and laboratory aviation research effort (Saleem, 2003).

Participants and Procedure

Three commercial pilots participated. One flew a visual approach to Roanoke Runway 33 and the other two flew an instrument approach using the ILS on Runway 33. Commercial pilots were recruited for the field study since FAA regulations disallow compensation of non-commercial pilots, § 61.113 (a) and (c) of Title 14, Code of Federal Regulations (U.S. Government, 2003). Each participant's prior flight experience was recorded (VFR, cross-country, IFR, simulated, and total hours).

After verifying that all data collection equipment was working properly, the pilot, safety pilot, and researcher took off from the Virginia Tech airport towards Roanoke, VA. The researcher was seated in the back seat and the safety pilot in the co-pilot's seat. The presence of a safety pilot was required by the Virginia Tech Institutional Review Board (IRB). The safety pilot was a licensed private pilot

certified for IFR piloting and his primary task during each flight was to watch for nearby traffic while the participants flew by instruments. He was also present to pilot the aircraft in the event the participant became incapacitated during the flight. Two pilots performed an instrument approach procedure to Roanoke Runway 33 and used a hood to simulate instrument meteorological conditions (IMC). The glide slope angle for the ILS at the Roanoke airport is 3-degrees, which is considered a standard approach angle. The third pilot performed a visual landing approach to the same runway.

The flight from the Virginia Tech Airport to the Roanoke airport lasted approximately 18 minutes for the visual approach and over 30 minutes for each of the instrument approaches using the ILS 33. Each participant's landing at the Roanoke airport was "touch and go" (a landing with an immediate takeoff) and then direct back to the Virginia Tech airport for debriefing. Average total participation time per participant including pre-flight preparation, the flight, and a debriefing session was two hours and 26 minutes.

Observations and Data Collection Instrumentation

Data collection consisted of video and audio recordings during the flight and of retrospective reports, field notes, and interviewing. GPS positional data was also collected to document the path of each flight. All data collection in the aircraft was observational, passive, and non-intrusive, consistent with naturalistic research norms. Except for a miniature video camera and GPS antenna, each component was secured to a board, which resided on the researcher's lap. A miniature pinhole video camera was securely mounted on the ceiling of the aircraft cockpit to obtain a complete video record of each participant's approach and landing to the airport.

Audio communications were captured directly from the cockpit intercom radio and relayed to the digital video creator. A special software template was created to capture the audio and video simultaneously; the audio communications were overlaid on the digital video and saved to the same file with time synchronization. Field notes from participant observation were recorded by the researcher (by hand) throughout the scenario as an additional means of qualitative data collection and to augment the video/audio recordings.

Another source of data included retrospective verbal reports. The video recordings were shown to participants promptly after returning to the Virginia Tech airport and they were asked to describe aloud what they were doing at the time of the recording. These retrospective verbal reports were audio recorded. Finally, a post scenario interview after each flight was conducted as additional means of collecting qualitative data. The questions were open-ended as the purpose was to acquire missing information (e.g., help the researcher understand some of the activities and communication that took place during the scenario).

Data Reduction and Analysis

All audio records from both the pilot/ATC communications and the retrospective reports during video playback were transcribed. Instances of pilot error, including any excerpts related to the event in question, were separated from the transcription record and classified as critical incidents by the researcher. All seg-

ments from the transcripts were coded in sequence so that any sorting of the segments into relevant groupings could be tracked back to the original transcripts.

Results and Discussion

We observed three critical incidents, one for each of the three flights (VFR01, IFR01, and IFR02), in which pilot error occurred. These events are summarized in Table 1. We discuss these instances and the present the relevant data from the transcripts from each of these three flights.

Table 1
Observed Error during Field Observation

Participant	Time Index	Transcript Code	Error
VFR01	6:55	A23-A24, A32-A36	Participant VFR01 initially failed to contact Roanoke Tower when instructed to do so by Roanoke Approach
IFR01	22:34	C127-C133	Participant IFR01 contacted Roanoke Tower and requested a landing for Runway 6 when intending to land on Runway 33
IFR02	16:11	E76, E78-E79	Participant IFR02 failed to attend to a course correction given by Roanoke Approach

Note: See Saleem (2003) for complete transcript records of each flight.

Traffic Avoidance

During the visual approach to Runway 33, participant VFR01 received the following communication from Roanoke Approach (Table 2):

Table 2
Hand-off of Cessna 61891 from Roanoke Approach to Roanoke Tower

Code	Time	Source	Communication
A23	6:55	RAD	Cessna eight-niner-one, enter left traffic Runway 33, and contact the tower one-one-eight point three.
A24	7:00	P	Left traffic for three-three, eight-niner-one.

Note: P = Pilot (Subject VFR01); RAD = Roanoke Approach and Departure

The pilot acknowledged the command to enter left traffic pattern for Runway 33 but did not acknowledge the command to contact tower. During the retrospective report, the pilot stated that he/she did not contact the tower at that point because he/she had not yet entered the left traffic pattern for Runway 33. However, circumstances suggest the pilot may have missed this command. The pilot failed to repeat the command to contact tower back to Roanoke Approach, as is routine. Further, when Roanoke Approach gives a command, it is meant to be followed promptly, unless otherwise specified. Therefore, it seems as if the command was afterward forgotten as over three minutes passed and then Roanoke Approach had to repeat the command as follows (Table 3):

Table 3

Second Attempt to Hand-off Cessna 61891 to Roanoke Tower

Code	Time	Source	Communication
A32	10:16	RAD	Cessna eight-niner-one contact the tower, one-one-eight-point-three.
A33	10:20	P	[Cessna?], tower, eight-point-three.
A34	10:24	P	Hello tower, Cessna eight-niner-one is with you, ah, descending down, [inaudible].
A35	10:28	Tower	Cessna six-one-eight-niner-one! Roanoke Tower, make a hard right turn for me heading about one-two-zero for right now please, vector away from departure traffic.
A36	10:35	P	One-two-zero, eight-niner-one.

Note: P = Pilot (Subject VFR01); RAD = Roanoke Approach and Departure; Tower = Roanoke Tower

After the pilot contacted the Roanoke Tower, the tower air traffic controller hurriedly instructed the pilot to make a hard right to avoid departing traffic. Had the pilot contacted tower earlier, he/she likely would have received navigational commands to avoid this potential conflict. If the repeated course correction did not arrive from ATC in time, the path of Cessna 61891 would have been in a potential and catastrophic collision conflict with the departing traffic.

Which Runway?

Another error occurred during the hand-off to Roanoke Tower from Roanoke Approach with participant IFR01 during the instrument approach. The pilot was on final approach for Runway 33, but mistakenly requested Runway 6 when making initial contact with the tower (it is common to fly the ILS for Runway 33 and then enter the airport's traffic pattern and land on a different runway). The initial communications with Roanoke Tower were as follows (Table 4):

Table 4

Pilot Communication to Tower of Runway Request

Code	Time	Source	Communication
C127	22:34	P	Calling tower Cessna eight-niner-one on the ILS three-three request a touch-and-go to Runway six if not then full stop to Runway six.
C128	22:43	SP	Three!
C129	22:44	P	Three-three, I'm sorry, Runway three-three.
C130	22:45	Tower	[inaudible] continue.
C131	22:47	P	I say again? Eight-niner-one.
C132	22:59	Tower	And, ahh, six-one-eight-niner-one, just ahh, continue [inaudible] you can plan a right base for Runway six when you're ready.
C133	23:04	P	Plan a right base six eight-niner-one.

P = Pilot (Subject IFR01); SP = Safety Pilot; Tower = Roanoke Tower

The pilot mistakenly requested Runway 6 and the safety pilot caught the error and yelled "three!" for Runway 33 to the pilot. The pilot tried to correct his mistake

with the tower but their communications interfered with each other as they communicated at the same time. The tower did not receive the correction and informed the pilot to plan a right base for Runway 6. Both pilots who flew the ILS approach to Runway 33 indicated in their retrospective reports that it is at this time during the flight that the pilot has the highest workload. During this time, the pilot is performing motor tasks as he/she pilots the aircraft, visual-spatial tasks by aligning the localizer and glide slope needles, and auditory tasks when listening for air traffic control commands and then repeating the commands back for error-checking. At this time, the aircraft is close to or over the Vinton non-directional beacon (NDB), so the pilot also is watching for the automatic direction finder (ADF) bearing indicator to swing as the aircraft passes over the NDB station at Vinton.

Approximately one minute later, the pilot contacted tower again, requested, and received clearance for Runway 33 rather than Runway 6. Had the safety pilot not been present to catch the error, it is possible that the pilot would have continued down and landed Runway 33, thinking he/she had asked for and received clearance for 33, while the tower was expecting him to land Runway 6. Potential for a catastrophic collision with another aircraft on Runway 33 would have been high. Such runway incursions, where aircraft use the incorrect runway do occur, most notably this year in Lexington, KY, where ATC had cleared the aircraft to take off from a particular runway, and the aircraft used a different, much shorter runway and crashed shortly after (National Transportation Safety Board, 2006).

Course Correction

The second participant who flew the instrument approach committed a less dramatic error when he/she failed to attend to a course correction given by Roanoke Approach at time index 16:11 (Table 5).

Table 5
Failure of Pilot to Respond to Specified Course Correction

Code	Time	Source	Communication
E76	16:11	RAD	Eight-niner-one, turn left heading of zero-five-zero.
			[pilot fails to respond]
E78	16:22	RAD	Number six-one-eight-niner-one, turn left heading of zero-four-zero now.
E79	16:26	P	Left heading zero-four-zero, eight-niner-one.

P = Pilot (Subject IFR02); RAD = Roanoke Approach and Departure

Not hearing a response from Cessna 61891, the air traffic controller would know that the pilot did not copy the command. The air traffic controller simply gave another course correction command 11 seconds later and this time, the pilot heard the command and responded. In the retrospective report, the participant confirmed that he/she did indeed miss the original course correction command. The video playback shows the safety pilot look over at the pilot, likely noticing that the pilot failed to attend the command. Had the error been a more critical one, as the one discussed for participant IFR01, the safety pilot would have informed the pilot of the error.

Design Implications

The literature categorizes different types of aviation errors. For example, Helmreich (2000) classified errors as “violation,” “procedural,” “communications,” “proficiently,” and “decision” errors. Another framework for errors categorizes them as decision errors (including procedural errors, poor choices, and problem solving errors), skill-based errors, and perceptual errors (Wiegmann, Shappell, Boquet, Detwiler, Holcomb, Faaborg, 2005). Weigmann et al. distinguished errors as occurring while aircrews are behaving within the rules and regulations of an organization and thus classify “violations” separately from errors. Violations, on the other hand, are behaviors that occur contrary to established ways of doing things, written and as practiced. The three critical incidents presented in this paper are best described as involving “communications errors”, by Helmreich’s categorization (Helmreich, 2000), where information exchanged from ATC to the pilot was not initially processed by the pilot (case 1 and 3) or the information exchanged was incorrect (case 2). These errors support practical design implications. Human-centered, aviation technologies may help minimize similar errors from occurring. Some of these technologies have been in place in commercial aviation, and similar systems could be modified for GA use, especially as computing technology advances to offer the benefits of these technologies to more pilots in smaller classes of aircraft. We present some examples of system redesign that may help address the incidents in each of the case studies presented. However, the following suggestions are meant to be examples of the more general need for human-centered design interventions to assist the pilot and minimize the potential for similar incidents from occurring, rather than specific recommendations necessarily derived from data. Further, any design recommendations would need to be tested in a follow-on research study to demonstrate a positive effect prior to implementation.

Communications filter. One potential design improvement would be to have the system filter out ATC communications that are not relevant to the pilot. The pilot receives most of the ATC communications occurring in the airspace as much of the traffic uses the same frequency. The more traffic in the airspace, the greater number of communications that are not relevant to a particular pilot. For example, consider the communications from the flight piloted by participant IFR02. ATC contacted participant IFR02 at time index 10:06 for a course correction and the participant immediately responded. However, a little over six minutes pass before ATC contacted participant IFR02 again at time index 16:11 and participant IFR02 missed the communication, after which ATC had to repeat the command. In the six minutes that transpired between contact with ATC, 13 communications were transmitted over the same frequency involving other aircraft, none of which were relevant to participant IFR02. Rather than the pilot attending to all communications and processing those that are relevant, the system could be designed to pass only relevant communications, thus reducing the burden on the pilot, and the potential for missing the communication as illustrated in this example. In such a communications filtering system, pilots should have the option to filter communications based on individual preference and only after they obtain training on the system so they can make an informed decision.

Displays to reduce burden on pilot working memory. There are several cases during the landing approach where the pilot must process and retain information

in working memory that is communicated from ATC. Such information includes specific course headings, altitudes, frequencies, and traffic information of nearby aircraft, amongst other piloting information. For example, the following are communications from ATC to participant IFR01 that contain a heading, altitude, frequency, and traffic information that must be remembered by the pilot (Table 6).

Table 6
Example Communications of Information from ATC the Pilot must Process

C86	15:38	RAD	Cessna eight-niner-one, turn left heading zero-six-zero. Suggested altitude for the ILS approach will be three-thousand-eight-hundred.
C124	22:16	RAD	Cessna six-one-eight-niner-one, you're at Vinton, contact tower one-one-eight-point-three and he will assign which runway for touch-and-go.
C136	23:27	Tower	Piedmont thirty-three-twenty-six, cleared for takeoff runway three-three, turn left heading two-five-zero.

RAD = Roanoke Approach and Departure; Tower = Roanoke Tower

This information could be “remembered” by the system rather than the pilot. Such a redesign would reduce pilot workload since the pilot would not need to immediately attend to incoming ATC communications for this information. Rather, the pilot could access this information from a display later if he/she was not able to process the information immediately. The pilot would still hear incoming ATC communications in real-time over the communication radio. However, the display would not only act as a redundant source of information in case the pilot failed to attend to the communication(s), but would also serve as a reference for error checking.

Auditory displays and traffic information systems. Traffic avoidance is another area of concern in the aviation domain and is underscored by the data collected from participant VFR01. Recall that the pilot in a VFR flight is responsible for maintaining clearance from nearby traffic. ATC will sometimes assist in VFR traffic separation, but the primary responsibility is the pilot's during VFR flights. The pilot receives positional information for surrounding traffic through ATC communications, as was the case for participant VFR01 (Table 7):

Table 7
Example ATC Communications with Positional Information of Nearby Traffic

A18	6:13	RAD	Blueridge four-zero-four, turn right heading zero-five-zero, direct Montebello.
A35	10:28	Tower	Cessna six-one-eight-niner-one! Roanoke Tower, make a hard right turn for me heading about one-two-zero for right now please, vector away from departure traffic.

RAD = Roanoke Approach and Departure; Tower = Roanoke Tower

The pilot listens to the communications from the air traffic controller and other pilots and develops an internal representation of surrounding traffic. Relevant parts of this traffic pattern are projected into the future for potential effects on his/her flight path. The pilot visually scans for traffic that is close to his/her aircraft until he/she has visual contact with that traffic.

As the pilot receives relevant traffic positional information from the communications, this information is transformed and represented internally in the pilot's working memory. In this case, the pilot would benefit from an auditory display of this traffic information. Three-dimensional auditory displays are one of the new technologies envisioned for future aviation systems. An auditory display is similar in concept to a visual display, except designed for the auditory sense. In other words, an auditory display uses sound to display information, often to enhance current user interfaces. For example, aural alerts of nearby traffic would allow the pilot to perceive the direction of nearby traffic, thus facilitating visual contact with traffic that may pose a potential collision hazard. Participant VFR01 had to rely on ATC for traffic avoidance of an approaching aircraft (see Table 3). An aural alert could have allowed participant VFR01 to pinpoint the relative direction of the other aircraft, potentially increasing his spatial situation awareness.

A traffic information system (TIS) could also assist a GA pilot with traffic avoidance. Aircraft can be equipped with new Mode S transponders, which are data-link capable. That is, they can receive data such as traffic information that is displayed in the cockpit with a TIS. A pilot can initiate a request for traffic information. Position, altitude, and distance information for nearby aircraft is then received through the data-link and displayed for the pilot.

Cost and Availability Considerations for Design Interventions

There is variability in cost and availability of any design interventions, including those suggested in this paper. Use of a communications filter is a relatively minor design modification and should have relatively low cost. Aviation displays, on the other hand, may carry a higher cost. There are off-the-shelf aviation displays, such as a TIS. A TIS-capable Mode S transponder (manufactured by Garmin or Honeywell for example) is needed to receive the information, and a multifunction or a multi-purpose GPS display is needed to depict the traffic graphically. Such a transponder and display could currently be purchased for less than 10,000 USD. If we consider the TIS a baseline for similar aviation displays, then more sophisticated systems that are capable of displaying additional information or have the capability of retaining information for the pilot to access when convenient would be more expensive. However, development of such displays should not be cost prohibitive for GA, as these systems do not involve new technologies; they simply involve taking advantage of current technologies to design enhanced displays. Auditory displays, on the other hand, may be currently cost-prohibitive, as these systems are relatively new and need further research and development. However, auditory displays have a promising potential to increase the spatial situation awareness of the pilot by providing positional information through the human auditory sense in addition to information already received through the visual sense. One barrier, and thus tradeoff, to use of newer aviation technologies is that there seems to be a considerable lag between research and actual design change, as an extensive and rigorous FAA certification plan is required for aviation design/redesign.

Enhancing Aviation System Design Using a Pilot-centered Approach

There are concerns with enhancing aviation technology without appropriate consideration of the potential negative impacts on the pilot. As aviation technology advances and displays become more advanced, designers must assess the

potential increased information load on the pilots (Schvaneveldt, Beringer, & Leard, 2002). In addition, more sophisticated technological systems can sometimes lead to poorer pilot performance in conditions where the validity of the information provided by the technology is uncertain (e.g., aviation decision support systems; (Vicente, 2003). Further, in cases where automation in the aircraft is increased, there is a danger of the pilot losing situation awareness. Thus, designers should include appropriate feedback mechanisms for the pilot to help maintain situation awareness. These are all important considerations to account for to ensure pilot-centered design.

Conclusion

We presented three critical incidents, involving communications errors, which occurred during field observations of pilots performing a landing approach in a GA aircraft to a mid-sized, controlled airport. One instance occurred during an attempted control handoff of the aircraft from one component of ATC (approach and departure) to another (tower). Another involved the pilot requesting the incorrect runway for a practice instrument approach. In the third case, an ATC command was missed by the pilot. In each of these instances, we showed examples of how modest aviation design enhancements may help the pilot in these situations, and potentially minimize similar scenarios from occurring. The design recommendations in this paper are some examples of user-centered approaches to redesign. However, further research study is needed to test the specific design recommendations provided in this paper to determine if they would have a positive effect, such as a simulation study. Other redesign interventions that may be more effective should also be considered and tested. Finally, design advancements for aviation systems must also consider potential negative impacts on the pilot, such as information overload and maintenance of situation awareness.

References

- Helmreich, R. L. (2000). On error management: lessons from aviation. *British Medical Journal*, 320 (7237), 781-785.
- Hollnagel, E., Woods, D. D. & Leveson, N. (Eds.). (2006). *Resilience engineering: concepts and precepts*. Brookfield, VT: Ashgate Publishing Company.
- National Transportation Safety Board (2006). *Preliminary report, Comair flight 5191*. (Report DCA06MA064). Washington, DC: NTSB. Retrieved from: <http://www.nts.gov/ntsb/GenPDF.asp?id=DCA06MA064&rpt=p>
- Roth, E. M. & Patterson, E. S. (2005). Using observational study as a tool for discovery: Uncovering cognitive and collaborative demands and adaptive strategies. In Montgomery, H., Lipshitz, R., & Brehmer, B. (Eds.) *How professionals make decisions*. (pp. 379 – 393). Mahwah, NJ: Lawrence Erlbaum Associates.
- Saleem, J. J. (2003). *Multi-method approach to understand pilot performance in a sociotechnical aviation system*. Unpublished doctoral dissertation, Virginia Tech (VPI&SU), Blacksburg, VA.
- Saleem J. J., & Kleiner, B. M. (2005). The effects of nighttime and deteriorating visual conditions on pilot performance, workload, and situation awareness in general aviation for both VFR and IFR approaches. *International Journal of Applied Aviation Studies*, 5(1), 107-120.

- Schvaneveldt R. W., Beringer, D. B. & Leard, T. M. (2002). *Evaluating aviation information systems: The role of information priorities*. (FAA Technical report). Retrieved from: <http://www.hf.faa.gov/Portal/techrptdetails.aspx?id=1773>
- Stanney, K. M. & Maxey, J. (1997). Socially centered design. In Gavriel Salvendy (Ed.) *Handbook of Human Factors, 2nd ed.* (pp. 637-656). New York: John Wiley & Sons.
- U.S. Government. (2003). *Title 14. In electronic code of federal regulations*. Retrieved from: <http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=%2Findex.tpl>
- Vicente, K. J. (2003). Less is (sometimes) more in cognitive engineering: the role of automation technology in improving patient safety. *Quality & Safety in Health Care, 12*, 291-294.
- Wiegmann, D. A., Shappell, S. A., Boquet, A., Detwiler, C., Holcomb, K., Faaborg, T. (2005). *Human error and general aviation accidents: A comprehensive, fine-grained analysis using HFACS*. (Technical Report AHFD-05-08/FAA-05-03). Savoy, Illinois: Aviation Human Factors Division, Institute of Aviation.

***Enhancing Life in the Hyper-Surveillance Mini-World
of a Space Station:
The Role of Situation Awareness, Communication,
and Reality TV in the Life of Astronauts***

William B. Rankin II

*Florida Memorial University
Miami Gardens, FL 33054*

and

John Cokley

*University of Queensland
Brisbane, 4072
Australia*

Abstract

This is the third article of a series entitled Astronauts as Audiences. In this article, we investigate the roles that situation awareness (SA), communications, and reality TV (including media communications) might have on the lives of astronauts in remote space communities. We examined primary data about astronauts' living and working environments, applicable theories of SA, communications, and reality TV (including media communications). We then surmised that the collective application of these roles might be a means of enhancing the lives of astronauts in remote space communities.

In January 2004, President George W. Bush proposed a reinvigoration of the United States' space program by proffering an initiative that would establish a long-term presence on the moon by 2020. This, in the President's view, would be the springboard for future travel to Mars and other space destinations. As part of the plan, the president advocated the resumption of the space shuttle program, which of course, has already occurred. He also advanced the development of a new space vehicle and the retirement of the current space shuttle fleet. An ambitious program, the President's announcement heralded the sentiment of many global citizens who still see space beyond planet Earth as our next frontier.

However, such a journey and such an outpost will require an extraordinary understanding of the factors and risks known to habituate space travel. Of those factors identified, none are as important as the psychological and the social requirements for space travel. In this article, we examined (a) primary data about astronauts' living and working environments, (b) applicable theories of SA, (c) communications, and (d) reality TV (including media communications).

Life as an Astronaut

Space as a Surveilled Workplace

For participants from developed, westernized countries, the career of astronaut is regarded as work, and one motivation for work is profit. Astronauts are also explorers and travelers, and explorers have a history of being paid for their labor, as do travelers have a history of being paid not only for the time they spend traveling but for the news and stories they bring back from those travels. Several astronauts including Edwin Aldrin (Apollo 11 lunar module pilot) and Gene Cernan (Apollo 17 commander) have produced books or articles about their time in space: in Aldrin's case, it was *Return to Earth* and in Cernan's, it was *Last Man on the Moon*.

The second man on the moon, Edwin Aldrin, famously submitted a travel expense claim to the National Aeronautics and Space Administration (NASA) officials, which detailed his claim as follows:

PAYEE'S NAME:	Col. Edwin E. Aldrin 00018
FROM:	Houston, Texas
TO:	Cape Kennedy, Fla. Moon Pacific Ocean
AMOUNT CLAIMED:	\$33.31 (Smith, 2005, p. 102)

Explorers are paid either for their skills, experience, know-how, and sense of adventure, or for their willingness to invest their own time and money in the quest for new lands, markets, or experiences. In the case of astronauts, this includes mission specialists and pilots - the crew - or payload specialists - the scientists who are employees. This option is the most common thought of when space travel is examined. However, with the emergence of SpaceShipOne, its pilot Mike Melvill and the subsequent arrangements with Virgin Galactic, this might not be the case for much longer. Space tourism is being promoted and marketed by Virgin Galactic as if it is real now, even though it acknowledges that the first commercial flights will probably not commence until 2008.

Astronauts also process information they observe and think about, just as any audience would. White (2002) noted, "...no one can go into space without filters..." and that "...to some extent we all create our own reality..." (p. 49):

The expressions of the experience are different, and therefore the transmission of the message varies widely ... (and) ... the specific social context of each flight, as well as the actual environment of space, is critical to the perception of the missions by astronauts and the public alike (White 2002, p. 50).

Primary Data about Astronauts' Living and Working Conditions

The physical situation of astronauts inside and outside a space station is relevant for establishing each astronaut's situation awareness. Astronauts' quarters are investigated, as are the wires and sensors that are hooked up to the body and what they measure, and extra-vehicular space suits. In this way, it is shown that astronauts do not have truly private moments. The only private moment for an astronaut is thinking, and even then, EEG measures brainwave patterns, indicating stress, pleasure, happiness, or other characteristics (Mizrahi & Pedley, 2004).

Astronaut Monitoring Systems

Astronauts are the most monitored people, both on and off our planet. From the skin out (and to some degree, under the skin as well), astronauts' daily lives are subject to complex and intensive monitoring by ground controllers and health officials. The clothing and equipment deployed onto the astronauts by their employers (NASA and other space agencies) also incorporates intensive monitoring and a particular living environment. Space suit manufacturer ILC Dover, which designed and developed garments for the Apollo and Skylab programs in the 1960s and 1970s and the existing Space Shuttle program, worked from a design brief which included a vacuum environment, temperature extremes from -180 to +277 degrees Fahrenheit and the ability to withstand the impact of micrometeoroids and orbital debris (ILC Dover, 1994). A close examination of the design of these suits reveals emerging technologies such as e-textiles and e-membranes, as well as a wide array of contact points for physiological surveillance.

This is planned to continue and increase in scope. One example of the depth and complexity of such research is the collaboration between NASA, Case Western Reserve University, Cleveland Clinic Foundation, University Hospitals, and The National Center for Microgravity Research – known wholly as the John Glenn Biomedical Engineering Consortium (GBEC). The GBEC is in the process of conducting fluid physics and sensor technology research to address problems in the areas of astronaut health, safety, and performance. Interdisciplinary research is also under way in the areas of biology, physical sciences, engineering, and medical research to develop techniques and equipment that will address health and safety issues from a distance (John Glenn Biomedical Engineering Consortium, 2005). Ten projects were selected for a three-year funding period, which will end fiscal year 2005 (John Glenn Biomedical Engineering Consortium, 2005). Future astronauts will likely be monitored and aided by some of these devices:

1. The therapeutic application of ultrasound, a high-frequency acoustic energy that prevents bone loss in microgravity conditions.
2. A portable device to measure human metabolic activity at a faster sampling rate that is presently available in space or on Earth.
3. An instrument for in vivo bioluminescent molecular imaging that could be used to create a new bio-dosimeter for measuring effects from ionizing radiation in space.
4. A head-mounted device, similar to night-vision goggles, that uses

noninvasive optical technologies to address problems as disparate as radiation damage that could cause cancer, blood glucose and its links to diabetes, and brain physiology.

5. A prototype instrumentation system to detect and report cardiac dysrhythmias remotely using wireless communications and a Web browser.
6. A micro miniature monitor for vital electrolyte and metabolite levels with adaptability, self-checking capability, and negligible power requirements.
7. A virtual-reality dual-track treadmill for NASA's ongoing development of exercise devices for space crews.
8. Biochip simulation capability that is tailored to space applications, incorporating the latest fluid physics, findings about capillary, multiphase flow, and surface science.
9. Fluorescent microscopy techniques to assess bone cell cultures and develop countermeasures against bone loss.
10. Miniature implantable Microsystems for the controlled release of medicines that are diffused into the body through tiny silicon nanomembranes. The pore size of the membranes can be designed to achieve different rates of release. (pp. 5-14)

Pay and Benefits

According to NASA (Astronaut Candidate Brochure, 2005), salaries for civilian Astronaut Candidates are based upon the Federal Government's General Schedule pay scale for grades GS-11 through GS-13. The grade is determined in accordance with each individual's academic achievements and experience. Currently a GS-11 starts at \$56,445.00 US per year and a GS-13 can earn up to \$104,581.00 US per year.

Other benefits include vacation and sick leave, retirement plan, and health and life insurance. No special benefits accrue astronaut's families. The benefits are the same as those offered to all other NASA employees. (Astronaut Candidate Brochure, 2005). Selected military personnel detailed to the Johnson Space Center remain in an active duty status for pay, benefits, leave, and other similar military matters. (Astronaut Candidate Brochure, 2005)

Spare Time and Relaxing

Russian cosmonaut Valery Ryumin once said, "all conditions necessary for murder are met if you shut two men in a cabin measuring five meters by six and leave them there for two months" (Baard, et al., 2003, p. 120). For most astronauts, a daily regimen of two hours of exercise is necessary to counteract the effects of extended space travel. Many, of course, consider exercise a form of recreation and relaxing, while at the same time reducing the stress of isolation.

The psychological challenge of long-term space flights includes monotony, boredom, lack of privacy, and gender tensions to name a few. Music, the sounds of planet Earth such as waterfalls, rain, urban street noise, etc., help to keep crews lively and happy. Audiovisual hookups with family and friends, family pictures, music and videos will likely consume much of the free time for future astronauts (Baard et al., 2003).

Teams training at isolated sites on Earth are used to help prepare astronauts for what to expect psychologically of themselves and their crewmates in terms of personality, problem solving and responses to stress. Expert systems to help crews through crises may play a role in future flights (Baard et al., 2003).

Communications technologies for colonization will be the soft technologies of group process: facilitation, alternative dispute resolution, and mediation. Some may experiment with terraforming as a form of relaxing (Baard et al., 2003).

In the future, space travelers will find their favorite games changed by the environment of space. “No matter what your favorite sport – or game, or dance, or art – may be, the widely varying environmental contexts that outer space presents will make almost all leisure pursuits extreme sports by 2100 ... Living in space will mean playing in space, and endless innovation in sports and the arts.” (p. 119)

Astronauts' Social Environment

Sells (1966) proposed A Model for the Social System for the Multiman Extended Duration Space Ship. In proposing such a model, Sells' intent, in general, was to address the issues of confinement, isolation, and stress associated with crew adaptability particularly during extended space missions. In particular, Sells was concerned “with group organization, structure, and interpersonal interaction of crew members in the environmental circumstances of a typical mission” (p. 1130).

In defining the conceptual model, Sells (1966) also discussed the “constraints expected in the space ship situation” (p. 1131) including the following:

1. A formal organization with prescribed responsibility patterns for the entire crew;
2. Crew composition characterized by an elite corps of highly selected, trained, and educated volunteer specialists, all extremely ego-involved in the program and mission;
3. Low organizational autonomy as a result of the NASA organizational and operational system and the affiliation of crew members with military and civilian career services;
4. Low formally prescribed status distance among crew members; and
5. High task demand and mutual dependence, under high levels of isolation, confinement, limitation or mobility, privacy, and environmental threat.

While Sells considered these characteristics of the space ship social system lacking, he offered them as a starting point in understanding behavioral issues associated with long-duration space flight and most importantly extending the specification of constraints so as to “formulate a set of principles of social structure and group behavior as hypotheses for further research” (pp. 1130-1131).

As such, Sells (1966) attempted to dimensionalize the situation of the space ship in terms of two conceptual elements, system structure (social situations) and

behavior patterns (modes of interaction). Within the context of system structure or social situation, Sells included the following factors: (a) goals and objectives (b) philosophy and value system, (c) personnel composition, (d) organization, (e) technology, (f) physical environment, and (g) temporal characteristics (p. 1132). Within the context of behavior patterns or modes of interaction, Sells included the following: [a] interpersonal behavior, [b] leadership style, [c] factors promoting or interfering with member motivation, and [e] other principally behavioral aspects of group functioning” (p. 1131). The importance of Sells work is its foundational status in that it set the tone for future studies, including the present one.

In 1972 a series of essays were published under the rubric Human Factors in Long Duration Space Flight. The report - an attempt to reconcile the “behavioral, psychological, and sociological factors of the ‘microsociety in a miniworld,” (Connors, Harrison, & Akins 1985, p. ch1-1) - was published by the Space Science Board of the National Academy of Sciences, whose “mission was to survey the scientific aspects of the human exploration of space” (The National Academies, Space Science Board, ¶ 1).

Connors, et al. (1985) attempted to build upon and broaden the earlier works of Sells (1966) and the Space Science Board (1972). As such, their findings were framed by three guiding assumptions: (a) psychological and social factors will become increasingly important determinants of the success or failure of future space missions, (b) it is essential to avoid premature commitment to a narrow perspective, and (c) some of the uncertainties regarding life in space can be reduced through careful and rigorous behavioral and social science research (ch. 1-2).

Their theoretical orientation was from a systems perspective, which assumes - in the case of long-duration spaceflight - “highly interdependent components (e.g., technical, biological, and social), such that variations in one component typically have repercussions in one or more of the others” (ch. 1-2).

Connors, et al. (1985) argued that while space missions had traditionally drawn astronauts from homogenous pools, that such missions as defined in this paper - of extended duration and of international cooperation - would draw its astronauts from diverse, heterogenic populations.

As individuals, astronauts have physiological and psychological needs including sex and sensory stimulation as well as higher-level needs as described by Maslow. On face value, the long-term viability of space communities relies on them being heterosexual groups to reduce the effects of aging and death; but this brings with it “tendencies towards individuation and hypoarousal” (The National Academy of Sciences - Space Science Board, 1972, p.165) which impact negatively on group effectiveness.

In a wider sense, a strong tendency towards social withdrawal and individuation has been identified in such communities to the extent that group viability is challenged and the remedy suggested is interpersonal communication, whether verbal or otherwise (Space Science Board 1972). For example, Connors, et al. (2005) noted that astronauts are removed from their home environments and

communities, at the same time as they are removed from the usual variety of social relationships, and placed in a miniworld or microcommunity (Connors et al., 2005). This microcommunity brings its own challenges, such as intellectual impairment, motivational decline, and social tensions (Connors et al., 2005).

Since one of the characteristics of long-duration space flight is unchanging environmental circumstances (The National Academies, Space Science Board, 1972), sensory and intellectual stimulus needs to be provided using external or, at least, artificial means, such as news from home or communication between individuals about matters other than task-oriented issues. Of these two, news from home is the most likely to succeed, since research indicates that continuous personal communication between individuals in confined environments tends to result in individuals becoming overexposed to each other, resulting in stress (The National Academies, Space Science Board, 1972).

It is also likely that attention will be required to satisfy higher order sociological needs, such as esteem and self-actualization (The National Academies, Space Science Board, 1972), given existing data that astronauts tend to be well-educated and highly motivated individuals (Cokley, Rankin, & Söhnlein, 2005). Such individuals also indulge in a process known as reality-testing during which they interact with their physical environment and look to others to confirm the perceptions they gather in this process. However, according to The National Academies, Space Science Board (1972), “the unreal world of a tiny space capsule far from nowhere, with a very limited number of other people with whom to compare notes, would seem to impede normal reality-testing” (p.166).

The importance of providing as much communications as possible with Earth for residents of space communities is emphasized (The National Academies, Space Science Board, 1972). However, the kind of communications to be emphasized is also relevant. Whether it is communications that involves mission data and operational issues, which are in superabundance on routine space missions (Connors, et al, 1985), but is not the same as communications that allows participants to reality-test, compare notes, or stay in touch, and the kind of communications identified in this article - surveillance - and the creation of digital doubles (Andrejevic, 2004; Connors, et al, 1985), seems to add very little to the communications experience of the participants in the space community.

Situation Awareness

There is a modicum of support for considering situation awareness (SA) as an appropriate area of enquiry for enhancing life in the hyper-surveillance mini-world of a space community. Sells (1966) posited the belief that a standard set of system structure characteristics would help identify, in some limited manner, the social requirements for extended space flight. While not termed “SA,” the Sells model assumed a level of awareness appropriate to environmental elements that make up a situation.

SA from the Human Communication Perspective

It is important to understand that, at least from the perspective of the symbolic interactionist, social situational awareness (SSA) is a foundation of human com-

munication (Wood, 1982). Moreover, while awareness may be understood from different perspectives, we assume that similarity in language suggests at least some similarity in concept. For this reason, we concluded that there are clear conceptual links between the human factors and human communications and that in reality they are imbedded within each other.

Early empirical studies in the area of social situational awareness began with Sells (1963a, 1963b), who advanced the notion that behavior is an interaction of inner and outer forces or “some form of mediated transaction between organism and environment” (p. 696) and thusly proffered the principle of interaction that is represented by the well-known equation $R = f(O \cdot E)$, where R refers to behavior, O refers to the organism, and E refers to the environment (1963a, p. 696; see also Cattell, 1963). This equation can be defined as the “physical-geographical, biological, social, and cultural factors that interact continuously with each other and with the individuals involved” (Magnusson, 1981a, p. 3). The outer forces—the environment—are the arena for the situation, which Magnusson (1981b) defined as “those parts of the total world that an individual can experience and interpret and does perceive and interpret as having reference to himself and his behavior” (p. 15).

The situation, as a concept, is rather hard to describe, primarily due to its complexity. For example, Sells (1963b) noted earlier attempts to identify and demystify environmental factors, situational dimensions, interrelated factors relevant to social situations, group dynamics, dimensions of group performance, situational factors, situational influences, and/or situational stimulus variables - those components or elements associated with behavior. In turn, he developed a taxonomy or outline of basics aspects of the total stimulus situation. Included in this taxonomy were (a) natural aspects of the environment; (b) man-made aspects of the environment; (c) description of task-problem, situation and setting; (d) external reference characteristics of the individual; and (e) individuals performing relative to others (pp. 9-13).

Developed primarily as a tool to measure stimulus situations, the real value of the scale is its integration of principle dimensions (e.g., weather, social institutions, biologically defined factors, group memberships, collective situations, etc.) to specifics types of situations. This, in effect, offers an opportunity to “clarify the effects of individual (inner) and situational (external) factors which account for significant variance in behavior” (Sells, 1963b, p. 13). This is extremely important when one considers the following:

The environmental influence on individual development and on actual behavior is always mediated via actual situations. It is in actual situations, with their physical-geographical and biological characteristics, that the cultural and social characteristics of the total environment are reflected and can be experienced by individuals. However, it is not just the information offered directly in specific situations that constitutes the environmental influence. Indirectly, great influence is also exercised by the cognitive structures, contents, affective tones, and coping strategies characteristic of an individual’s conceptions of the total world and formed in earlier confrontations with various environments. In some sense, past environments are also present. In addition, the norms, values, goals,

paths, and other factors that determine the behavior of individuals in a given situation are embedded in and determined by the social and cultural environment at more distal levels. (Magnusson, 1981a, p. 3)

Another factor that makes situational understanding complex is the fact that definitions of the environment must consider the “conceptual distinction between (1) the environment ‘as it is’ and (2) the environment ‘as it is perceived,’ construed, and represented in the mind of an individual who is appearing and acting in it on a certain occasion.” (Magnusson, 1981a, p. 3) Thus, Magnusson (1981b) defined these two environments in terms of the actual situation and the perceived situation.

Magnusson (1981b) defined the actual situation as follows: “In physical and biological terms, a situation can be rather strictly defined as that part of the total environment that is available for sensory perception for a certain amount of time” (p. 14). In addition, Magnusson stated, “to the physical and biological properties of places . . . are attached sociocultural factors - norms, rules, roles, etc. - that contribute to a complete definition of an actual situation” (p. 14). Magnusson defined the perceived situation “as an actual situation as it is perceived, interpreted, and assigned meaning, or in other words, as it is construed by and represented in the mind of a participant” (p. 14).

For purposes of this investigation, the perceived situation is most important. Magnusson (1974) made the following observations concerning the individual and the situation:

1. The characteristic of a situation determines to some degree the behavior of an individual in the situation.
2. The kind and degree of influence on behavior differ from situation to situation dependent on the character of the situation; some situations are weak—some are strong, some situations are ambiguous—some are unambiguous, some situations are relevant - some are irrelevant.
3. The degree to which a situation determines the behavior of an individual differs from individual to individual.
4. The kind and degree of influence of a specific situation on individual behavior is dependent on the meaning or significance that the individual gives to the situation. Individuals may give different meanings to one and the same situation. (p. 125)

When one considers that situations are ever evolving, it becomes critical to define, for purposes of assessment and/or evaluation, particular incidents or occurrences within the continuum of time and space. For purposes of analysis, a particular event is typically the focus. Thus, in the context of the environment over time the momentary situation becomes important. Magnusson (1981b) defined the momentary situation “as the interface of the ‘vertical’ distal-proximal dimension and the ‘horizontal’ time dimension” (p. 17).

The importance of the momentary situation rests upon the contextual basis of the experience through which the individual senses, within the environment, those stimuli or events that produce a given behavioral action or reaction. Thus, according to Magnusson (1981b), the momentary situation might be viewed, at

least from a research perspective, in terms of three spaces - (a) as a space of action, (b) as a space of observation, with the actor as the observer, and (c) as a space of observation, with another person or other persons as observer(s).

The practical implication of a space of action and a space of observation in terms of the momentary situation is the ability to analyze situations more accurately. By taking snapshots of an evolving scenario or event, particular environmental elements are exposed. As Magnusson (1981b) noted:

Situations present, at different levels of specification, the information that we handle, and they offer us the necessary feedback for building valid conceptions of the outer world as a basis for valid predictions about what will happen and what will be the outcome of our own behaviors. (p. 9)

Communications

What is communication? Farace, Monge, and Russell (1977) defined communication from a structural-functional perspective as “the exchange of symbols that are commonly shared by the individuals involved, and which evoke quite similar symbol-referent relationships in each individual” (p. 26). Wood (1982), based on a symbolic interactionist perspective, defined communication as “a dynamic, systemic process in which communicators construct personal meanings through their symbolic interactions” (p. 20). Regardless of the definition, symbols are considered important features of communication.

According to Browne, Fishwick, and Browne (1990) “symbols surround and engulf us as emblems, tokens, signs, images. They are part of the hidden language that makes the spoken language possible. Civilization depends on symbols to supply meaning” (p. 1). Wood (1982) made the following observation: “The symbolic interactionist view of a situation calls attention to the role of language in defining situations and in directing our actions within them. It is through symbols that we indicate to ourselves certain external factors and our own internal states about those factors” (p. 40).

Earlier we considered Magnusson’s (1981b) definition of situation as “those parts of the total world that an individual can experience and interpret and does perceive and interpret as having reference to himself and his behavior” (p. 15). Wood (1982) offered the following explanation: “situation refers to interactions between external and internal factors that an individual symbolically indicates to himself or herself in order to organize and evaluate experience and to direct personal behavior” (p. 41). Further, according to Wood, “situations are constituted as individuals select, order, and interpret disparate external circumstances to create patterns that are meaningful to them and that form foundations for their own behaviors” (p. 39).

From the communications perspective, there are four types of situations: (a) intrapersonal communication situations, (b) interpersonal communication situations, (c) public communications situations, and mass communication situations. There are also three dimensions of communicative situations: (a) the dimension of purpose, (b) the dimension of environment, and (c) the dimension of persons and relationships.

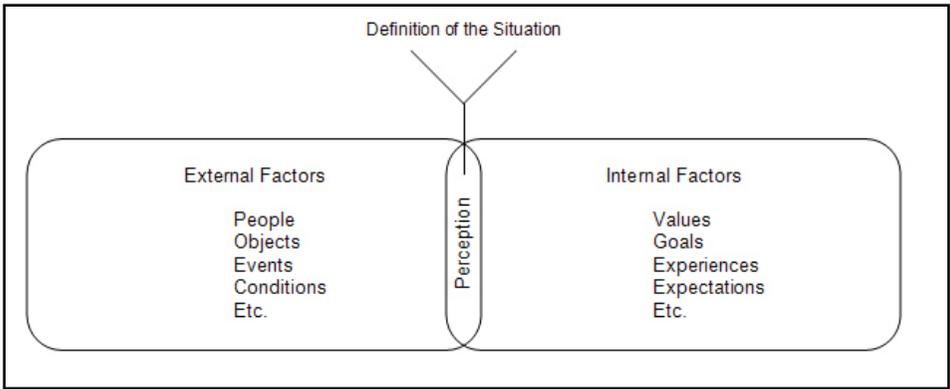


Figure 1. Model of the Individual's Definition of the Situation (Wood, 1982, p. 41).

Reality TV: Creating the Stimulus

The flood of reality TV - or what Lewis (2004) called factual entertainment (p. 288) - has literally drowned an all-consuming, media-frenzied populace, particularly in the United States (Andrejevic, 2004; Bennett, 2005; Frisby, 2004; Poniewozik, 2004). Otherwise, unknown folks become instant celebrities doing things most of us would never do. Conversely, the rich and famous - the stars and celebrities of yore - unmask their royalty to expose the simple, mundane, ordinary existence they purportedly live. We find ourselves uncontrollably gripped by the game show, the talk show, and the reality show—reality TV. What then is reality TV?

Ouellette and Murray (2004) defined “reality TV as an unabashedly commercial genre united less by aesthetic rules or certainties than by the fusion of popular entertainment with a self-conscious claim to the discourse of the real” (p. 2). Andrejevic (2004), on the other hand, defined reality TV - very narrowly - as “unscripted entertainment reliant on willing submission to comprehensive monitoring of the rhythm and events of daily life” (p. 64). “It documents an evolving situation as it happens” (p. 12). Frisby (2004) defined reality TV as “a genre of programming in which the everyday routines of ‘real life’ people (as opposed to fictional characters played by actors) are followed closely by cameras” (p. 51).

Central to all of these definitions is the notion of surveillance. Surveillance, a mediated spectacle according to Andrejevic (2004), is the essential apparatus that exposes the otherwise unknown. Nevertheless, surveillance is more than just an expose; it is a journey. This journey propels us to a different place, a different time, and even a different dimension. It gives us a perceptible glance into another's mysterious world, demystifying at least some aspects of another reality. The question is; nevertheless, is there utility to this phenomenon?

There is general support for the contention that media provides many positive features. For example, Frisby (2004) noted the following: “Researchers frequently refer to at least six gratifications of media use: information (also known as surveil-

lance or knowledge), escape, passing time, entertainment, social viewing/status enhancement, and relaxation” (p. 52). Beyond these gratifications are the very specific qualities attributed to reality TV. For example, Frisby also posited that reality TV serves as a medium for comparison. Known as social comparison theory, “individuals compare themselves with others for a variety of reasons, including to: determine relative standing on an issues or related ability; emulate behaviors; determine norms; lift spirits or feel better about life and personal situations; and evaluate emotions, personality, and self-worth” (pp. 52-53).

Poniewozik (2004) made an interesting claim. Despite our disdain, contempt, and even hatred for the relatively new genre, we still find ourselves living vicariously “through a bunch or 15-minute-fame seekers” (p. 12). Why? In many cases, it tells, “the quintessentially American story of ambition and desperation and shrinking options” (p. 16). And, as Frisby (2004) noted, “Reality TV allows audiences to laugh, cry, and live vicariously through so-called every day, ordinary people who have opportunities to experience things that, until the moment they are broadcast, most individuals only dream about” (p. 53).

Therefore, taking from the phenomenon of reality TV and its attendant media surveillance, one may glean a broadened understanding of and novel techniques to deal with such issues as astronauts’ lack of, loss of, or faulty situation awareness (their perceptions of reality), their psychosocial needs, their emotional needs, and even their fantasies. Remembering that for purposes of this article, the astronaut is the audience, how might space denizens be supported and their lives enhanced while in the space environment through the medium of reality TV?

Andrejevic (2004) observed that digital media is well suited for the creation of one’s own reality programming. As such, it allows individuals to do the impossible: seize control of their own media products. In addition, according to Andrejevic, this individualized control is a source of personal empowerment that allows for unprecedented manipulation and exploitation of mass media, particularly digital formats.

Questions in the opening section -- such as “Where am I?” or “Where is everybody else in relation to me?” and “Who is responsible for what?” “Who is in charge?” “Who is going to make what decision?” and “Who can perform what actions?” -- suggest that a detailed understanding of the space environment is integral to the most beneficial operation of that environment. This environment is analogous to a fishbowl with wires. A feature known as “presence awareness” (Andrejevic, 2004) is created by the various visual, auditory and electronic surveillance of astronauts, including telemetry. This allows engineers, doctors and others at Mission Control to bring into being what Andrejevic and others (2004:77) identify as a “data (or digital) double” of each astronaut, and the various interactions on board and of the community as a whole.

Existing paradigms allow communications on long-duration spaceflights to be characterized as direct interpersonal or mediated (Connors, et al., 2005, p. 194), but it is suggested that the kinds of mediated communications involved in creating and monitoring digital doubles lies somewhere between these extremes, since

even direct interpersonal communications involving space communities are almost all mediated by technology.

Discussion

Andrejevic (2004) suggested that the more an audience is encouraged to customize the media information it receives, the more the audience is open to surveillance and the less free it becomes. However, in the case of astronaut communities, surveillance is already maximized; suggesting that customization might be a positive side-effect of this level of surveillance, and that freedom to choose is, ironically, enhanced. Thus, this study suggested that heightened situation awareness is positively related to development of audience characteristics within space communities and could enhance individuals and communities' receptiveness to news from home.

The audience characteristics so formed are useful not just operationally but socially, and form part of the experience which helps each individual create not only their own identity but the social identity of each community within which they live, as well as helping to form the social identity of space and space travel among populations remaining on Earth. This is supported by the assertion that the creation and maintenance of individual and community identity is fundamentally a communications task and requires an understanding of communications theories and models, combined with situation awareness, in order to function appropriately.

Conclusion

In this article, we examined (a) primary data about astronauts' living and working environments, (b) applicable theories of SA, (c) communications, and (d) reality TV (including media communications).

As has been noted in this study, astronauts include a particular combination of four characteristics: explorer, traveler, storyteller (i.e. journalist), and audience. It has also been shown that astronauts – along with, and as part of, their operational exploring role – observe, process information and think about it, just as any audience would, that “no one can go into space without filters” and that “to some extent we all create our own reality” (White, 2002, p. 49). With the enhanced ability to study, understand, and predict those filters, which is suggested in this article, the future for space communities takes on a brighter and more positive aspect.

To achieve the ability to study understand and predict those filters suggested in this article, the space agencies of the world will need to work together to construct a vast new array of information highways; an interplanetary and eventually an intergalactic internet network of communications satellites orbiting the Earth, Moon and distant planets. This new internet will allow space explorers of the future to enhance their collective inter-personal and inter-group communications within, between, and among space communities. As a result, the future for space communities takes on an even brighter and more positive aspect.

References

- Andrejevic, M. (2004). *Reality TV: The work of being watched*. Lanham, MD: Rowman & Littlefield Publishers, Inc.
- Astronaut Candidate Brochure (2005). *Pay and benefits*. Retrieved July 27, 2005, from <http://astronaut.candidates.gov>
- Baard, E., Grierson, B., Sauls, B., Fenella, S., Schultz, W., & Winters, J. (2003). *Space 2100*. New York: Time Inc.
- Bennett, W. L. (2005). Beyond pseudoevents: Election news as reality tv. *American Behavioral Scientist*, 49, 364-378.
- Browne, R. B., Fishwick, M. W., & Browne, K. O. (Eds.). (1990). *Dominant symbols in popular culture*. Bowling Green, OH: Bowling Green State University Popular Press.
- Cattell, R. B. (1963). Formulating the environmental situation and its perception in behavior theory. In S. B. Sells (Ed.), *Stimulus determinants of behavior* (pp. 46-75). New York: The Ronald Press Company.
- Cokley, J., Rankin, W. B., II, & Söhnlein, G. (2005). Astronauts as audiences: Characteristics of the first space communities. *International Journal of Applied Aviation Studies*, 5(1), 167- 181.
- Connors, M. M., Harrison, A. A., & Akins, F. R. (1985). *Living aloft: Human requirements for extended spaceflight*. Retrieved July 11, 2006, from National Aeronautics and Space Administration (NASA), NASA History Office Web site: <http://www.hq.nasa.gov/office/pao/History/SP-483/cover.htm>
- Farace, R. V., Monge, P. R., & Russell, H. M. (1977). *Communicating and organizing*. Reading, MA: Addison-Wesley Publishing Company.
- Frisby, C. M. (2004, September). Getting real with reality tv. *USA Today*, 133, 50-54.
- Hayashi, M., Huemer, M. S., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005). Effects of the space shuttle cockpit avionics upgrade on crewmember performance and situation awareness. Proceedings of the *Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 54-58). Santa Monica, CA: The Human Factors and Ergonomics Society.
- ILC Dover (1994) *The shuttle space suit assembly*. Also ILC Dover website information, <http://www.ilcdover.com>
- John Glenn Biomedical Engineering Consortium (2005). *Helping astronauts, helping people on Earth*. Retrieved July 27, 2005, from <http://microgravity.grc.nasa.gov>
- Lewis, J. (2004). The meaning of real life. In S. Murray & L. Ouellette (Eds.), *Reality TV: Remaking television culture* (pp. 288-302). New York: New York University Press.
- Magnusson, D. (1974). The individual in the situation: Some studies on individuals' perception of situations. *Studia Psychologica*, XVI, 124-132.
- Magnusson, D. (1981a). Problems in environmental analyses—An introduction. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactional perspective* (pp. 3-7). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Magnusson, D. (1981b). Wanted: A psychology of situations. In D. Magnusson (Ed.), *Toward a Psychology of Situations: An Interactional Perspective* (pp. 9-32). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Mizrahi, E.M. and Pedley, T.A. (2004) In memoriam Peter Kellaway PhD, *Epilepsia*, Volume 45 (93), January 2004
- Ouellette, L., & Murray, S. (2004). Introduction. In S. Murray & L. Ouellette (Eds.), *Reality TV: Remaking television culture* (pp. 1-15). New York: New York University Press.

- Poniewozik, J. (2004). Reality TV has a positive influence on society. In K. F. Balkin (Ed.), *Reality TV* (pp. 12-16). San Diego: Greenhaven Press.
- Sells, S. B. (1963a). An interactionist looks at the environment. *American Psychologist*, 18, 696-702.
- Sells, S. B. (1963b). Dimensions of stimulus situations which account for behavior variance. In S. B. Sells (Ed.), *Stimulus determinants of behavior* (pp. 3-15). New York: The Ronald Press Company.
- Sells, S. B. (1966). A model for the social system for the multiman extended duration space ship. *Aerospace Medicine*, 37, 1130-1135.
- Smith, A. (2005). *Moondust: In search of the men who fell to earth*. London: Bloomsbury.
- The National Academy of Sciences - Space Science Board. (1972). *Human factors in long-duration space flight*. Washington, D.C.: National Academic of Sciences
- White, M. (2002). Television, therapy, and the social subject; or, the TV therapy machine. In J. Friedman (Ed.), *Reality squared: Televisual discourse on the real* (pp. 313-322). New Brunswick, NJ: Rutgers University Press.
- Wood, J. T. (1982). *Human communication: A symbolic interactionist perspective*. New York: Holt, Rinehart and Winston.

Transfer of Training from Flight Training Devices to Flight for Ab-Initio Pilots

Nickolas D. Macchiarella,

Pamela K. Arban

and

Shawn M. Doherty

*Embry-Riddle Aeronautical University
600 S. Clyde Morris Blvd.
Daytona Beach, FL 32114-3900*

Abstract

The application of flight simulation to meet pilot training needs continues to evolve. Flight simulations built with powerful and inexpensive computers are making high fidelity simulation available as a medium for training ab-initio pilots at Pilot Schools and Training Centers. The researchers conducted an 18-month study that applied an experimental flight-training curriculum comprised of 60% flight training device (FTD) flight and 40% airplane flight to certify Private Pilots under Federal Aviation Regulation (FAR) Part 142. The results from the research provided data to ascertain the effective transfer for each flight-training task. Ab-initio student pilots practiced each task to standard in an FTD prior to training in an actual airplane. The researchers measured a significant degree of effective transfer for the majority of flight tasks examined.

Introduction

As flight simulations increase in fidelity and decrease in relative cost, the possible applications of simulation for training necessitates continued investigation. Flight training devices (FTD) frequently include high levels of fidelity for aerodynamic modeling. The Federal Aviation Administration (FAA) defines an FTD as, a full scale replica of an airplane's instruments, equipment, panels, and controls in an open flight deck area or an enclosed airplane cockpit, including the assemblage of equipment and computer software programs necessary to represent the airplane in ground and flight condi-

tions to the extent of the systems installed in the device; does not require a force (motion) cueing or visual system; is found to meet the criteria outlined in this AC for a specific flight training device level; and in which any flight training event or flight checking event is accomplished (Federal Aviation Administration, 1992, p. 3).

Recently developed FTDs often include visual systems, force cueing, and aerodynamic modeling characteristics that were not readily available when the FAA first defined and then prescribed how these nonmotion-based flight simulators could be used for pilot training. The purpose of this study was to examine the transfer of training (ToT) from specific recently developed FTDs for ab-initio pilots who trained with a hybrid curriculum of simulated and actual flight.

Starting August 2005, Embry-Riddle Aeronautical University (ERAU)'s Daytona Beach campus conducted an 18-month research project as part of its effort to optimize the application of training that heavily integrates simulated flight during the training of ab-initio student pilots. This longitudinal study followed the performance of participant pilots from a novice condition to FAA certification as a Private Pilot. The transition to powerful personal computer (PC) systems used to drive FTDs is enabling higher levels of fidelity at lower costs while accurately modeling specific aspects of flight. The researchers examined the skill transfer from a Frasca 172 FTD to single engine airplanes for ab-initio student pilots. The Frasca 172 FTD is an FAA qualified Level 6 FTD with a 220-degree wraparound visual system and enhanced aerodynamic modeling that includes non-linear dynamic coefficients, accurate p-factor, slow flight, stalls, left turning tendencies, and force cueing (Anderson & Macchiarella, 2004). This study differed from previous skill transfer studies due to its application of a hybrid curriculum based primarily in simulation and the degree of simulation use approved by the FAA to certify pilots under Federal Aviation Regulation (FAR) Part 142. The research used Flight Training Devices for 60% of the hybrid curriculum's training while airplanes were used for the remaining 40%.

Transfer of Training

Evidence exists indicating that flight training in simulators can yield a positive transfer to performance in real flight (Hays, Jacobs, Prince, & Salas, 1992; Rantanen & Talleur, 2005; Stewart, Dohme, & Nullmeyer, 2000; Waag, 1981). For example, Stewart, Dohme, and Nullmeyer (2000) replaced 7.8 hours of training in the aircraft with nine hours of training in a relatively low fidelity simulator for a group of ab-initio pilots. The measures for this experiment included whether students were set back or eliminated from the training program. The simulation-based training resulted in the experimental group students achieving standardized performance without being set back or eliminated. These findings led Stewart et al., (2000) to conclude that simulated flight had utility for ab-initio flight training.

Using a meta-analysis Rantanen and Talleur (2005) found few differences between transfer of training (ToT) studies that used simulation to train instrument tasks and those that trained visual tasks. This suggested that the procedural aspects of visual flight could be effectively trained in simulation. They also found that training in a conventional simulator without a visual system becomes less cost effective as training extends beyond ten hours. However, simulators with

some type of visual system offer new cues and may be potentially cost effective beyond ten hours, although the nature of their effectiveness is not well documented.

Additional studies examined the application of simulated flight to train pilots that ranged in skill levels from ab-initio student pilots to experienced airline transport pilots (ATP) (Brown, Cardullo, & Sinacori, 1989; Go, Bürki-Cohen, & Soja, 2000; Jacobs & Roscoe, 1975; Waag, 1981; White & Rodchenko, 1999). However, these studies did not specifically address the use of modern high fidelity FTDs to meet the training needs of ab-initio student pilots.

Although previous studies demonstrated the effectiveness of simulation for flight training, questions remained regarding how effective simulation is for training initial flight skills in ab-initio pilots as findings in prior work have generated mixed results (Rantanen & Talleur, 2005). The need for further research examining the effect of FTDs on ab-initio pilot training remains open for examination.

Measurement of Transfer

The most common method of measuring the degree of skill transfer between simulation to the aircraft in order to determine simulation effectiveness is ToT (Roscoe & Williges, 1980). The concept of ToT is derived from learning theory. Researchers have shown that learning and skill acquisition can be transferred from one setting to another similar setting (Gerathewohl, Mohler, & Siegel, 1969). Existing skills can either help or hinder the learning and development of new skills. When pre-existing skills have a positive effect on the development of a new skill, the change in skill is referred to as positive transfer. Conversely, hindrance of new skill acquisition by pre-existing skills is called negative transfer (Patrick, 2003; Roscoe & Williges, 1980). The degree of positive or negative transfer can be measured by a transfer effectiveness ratio (TER) (Roscoe & Williges, 1980).

Calculating the TER requires counting the practice iterations for a task until experimental and control group participants achieve prescribed levels of proficiency in their respective training program. A TER is calculated by subtracting the number of iterations of a task in the aircraft performed by the experimental group from the number of iterations for that task in the aircraft for the control group. This resultant number is subsequently divided by the number of iterations in the simulator (i.e., an FTD) by the experimental group (see Figure 1) (Roscoe & Williges, 1980). Higher TERs indicate greater transfer from simulation to actual airplane flight (e.g., a TER of 1.0 indicates a higher level of transfer than a lower TER like 0.4). A TER of one indicates that for every iteration in the FTD one iteration is saved in the airplane. All positive ratios demonstrate savings in airplane flight for the experimental group.

$\text{TER} = \frac{C - E}{E_{(FTD)}}$
C is Control Group Task Iterations in an actual airplane E is Experimental Group Task Iterations an actual airplane $E_{(FTD)}$ is Experimental Group Task Iterations in an FTD

Figure 1. TER Formula and Definitions of Terms

As an example, Stewart et al. (2000) pre-trained ten pilots in simulated flight for eight flight tasks. They recorded the number of iterations necessary to achieve standard in the aircraft following the pre-training in simulation. A control group of 21 pilots received no simulated flight training. These researchers found that for all eight maneuvers pre-trained, there was a positive ToT from simulated flight to aircraft flight. The overall TER for the eight tasks was 0.55. This number indicated that each iteration practiced in simulated flight saved approximately one-half iteration during aircraft flight training. Similar transfer effectiveness ratios were anticipated for the current study at ERAU using an FTD with greater fidelity and a curriculum that is tailored specifically for the incorporation of simulation. The researchers hypothesize the ab-initio pilots participating in flight training integrating Frasca 172 FTDs and real flight will meet training standards with significant TERs.

Methods

Participants

This study used 38 participants: 18 were in an all-flight control group, and 20 were in an experimental group that used the hybrid FTD and airplane flight curriculum. The number of participants for each group was selected based on a previously conducted power analysis that indicated that 18 participants would be adequate for an in-study power of .80. All participants were ab-initio student pilots with a trivial amount of previous flight training (mean of 0.24 flight hours). The mean age of the all-flight control group was 18.5 and the mean age of the experimental group was 18. The all-flight control group contained 14 males and 4 females and the experimental group contained 15 males and 5 females. The participants were seeking a Bachelor of Science in Aeronautical Science at ERAU. They were regularly enrolled students seeking credits. Flight cost for study participants were normalized to the university's regular flight costs. Each participant possessed, as a minimum, a current Class III Medical Certificate.

Apparatus

This research used aircraft and FTDs obtained from the university's regular training fleet. Flight instructors used the Cessna C-172S "Skyhawk" for 100% of the control group's flight training and 40% of the experimental group's flight training. The C-172S was equipped with NAV II Avionics that includes traditional round dial instrumentation, Garmin 430, global positioning system (GPS), VOR, and DME. Additionally, the Frasca 172 FTD was used for the bulk (i.e., 60%) of the experimental group's curriculum. A Level 6 FTD is defined as a non-motion training aid that is aircraft specific (Federal Aviation Administration, 1992). The device used at ERAU was qualified as a Level 6 Flight Training Device. This device was further enhanced to handle the high angle of attack envelope necessary to train ab-initio pilots. The new aerodynamic models necessary to achieve the desired fidelity were longitudinal and lateral-directional propeller destabilizing effects, longitudinal and lateral-directional gyroscopic effects, p-factor, stall model, and an asymmetric wing lift (spin) model. The researchers at ERAU referred to these FTD as being Level 6 Plus to reflect these enhancements. These FTDs are embedded within an actual Cessna C-172S cockpit from the front of the airplane to just behind the pilot seats. From the firewall forward, the FTDs house the flight control loading equipment. Behind the pilot seats is an instructor's station with a

computer workstation to monitor and control the simulation. The visual system provides a 220-degree out-of-the-cockpit view of the flight environment. Air vents in the cockpit blow air on the pilot to replicate the cabin airflow levels experienced in flight. Aural cues change dependent on RPM settings, flap movements, stall warning, airspeed, and engine power. The radio and intercom systems functionality match actual radio and intercom systems in a C-172S and have the capability of being networked with other FTDs for a fleet-wide simulation. All of the Frasca 172 FTDs are equipped with Global Positioning System (GPS) and Instrument Landing System (ILS) navigation capabilities. The aerodynamic modeling is based upon comprehensive flight test data collected at ERAU of a full flight regime to include all aspects of slow flight and stall performance (Anderson & Macchiarella, 2004). This configuration is currently commercially available from Frasca International, Inc.

Flight Training Curricula

Two separate curricula were approved by the FAA to conduct this study. The curricula were structured to sequence tasks in the same manner for both groups. Variations in the curricula between the control and experimental group were minimized. Very little research has been done concerning when and how simulation should be integrated into a flight-training curriculum (Champney, Milham, Carroll, Stanney, & Cohn, 2006). The curriculum selected for training the experimental group (Embry-Riddle Aeronautical University, 2005a) was designed to sequence flight tasks first in the FTD with the goal of obtaining FAA prescribed criterion prior to aircraft flight. In cases where the Practical Test Standard (PTS) was ambiguous with regard to measurable task completion, researchers applied a criteria derived from the ERAU Standard Operating Procedures (SOP) (Embry-Riddle Aeronautical University, 2003) (e.g., Preflight Inspection and Cockpit Management). The experimental curriculum contained 60% simulated flight and 40% airplane flight for a total of 69.7 hours of flight training. Students successfully training with this curriculum had approximately 28 hours of flight in the real aircraft. The control group's curriculum was comprised of 100% aircraft flight.

Data Collection

The FAA Private Pilot Practical Test Standards for Airplane (SEL, MEL, SES, MES) (Federal Aviation Administration, 2002) served as the source for task criteria. Instructor pilots collected data by recording task iterations on paper forms for each participant during training flights. Iterations included any attempt by the student to perform a PTS task including successful completion of the task to standard. Data were collected in the same manner during training with the airplane and FTD. Researchers chose to utilize the tasks from the PTS as the measurement of pilot performance due to the PTS's regulatory authority in the certification of pilots. All FAA certified flight instructors and pilot examiners must comply with these standards when conducting practical tests that come at the end of the Private Pilot certification course. The PTS standards specify the Areas of Operation for which students must show competency before receiving a Private Pilot certificate (Federal Aviation Administration, 2002). Researchers listed each task from these respective Areas of Operation on a paper data collection form referred to as an iteration slip. On the iteration slip, instructor pilots recorded each iteration of PTS tasks attempted and whether the iteration was successfully completed. Iteration slips were bound and placed on kneeboards to accommodate the less-than-

optimal data collection conditions that instructor pilots experience during flight training. The data collection device listed all tasks on the front side of the iteration slip in large print (Beaubien, 2004).

As part of the effort to enhance reliability, the instructor pilots received data collection training to standardize the collection procedures. The standardization occurred through an initial 3-hour training period and subsequently reinforced with a monthly review of procedures. The instructor pilot training addressed the PTS standards and necessity of adherence to the curricula approved by the Orlando Flight Safety District Office (FSDO)-15 (Embry-Riddle Aeronautical University, 2005a, Embry-Riddle Aeronautical University, 2005b).

Design

This experiment was a two group between subjects design. The independent variable was the training platform. The control group's condition included full flight in the C-172S with no FTD exposure. The experimental group's condition contained C-172S flight and FTD flight. There were 34 dependent variables. These 34 dependent variables were represented by the number of iterations necessary to achieve the PTS standards for 34 tasks associated with Private Pilot certification.

Procedure

The university institutional review board examined all procedures and approved the study. Researchers carefully followed the approved procedures. Participants were pre-briefed and randomly assigned to a group before signing an informed consent form. As students registered to participate in the study the university registrar randomly divided the students into groups and, when all participants were assigned, sent a list of participants for each group to the researchers. Each group had a slightly different informed consent form to account for the different benefits available to each group for participation in the study. Researchers briefed members of both groups on their respective curriculum. Each participant filled out biographical data including contact information and number of flight hours. All participants were screened to ensure they had no more than a trivial amount of prior flight experience (i.e., < 1 hour).

Participants received the same academic ground training as the general population of Aeronautical Science students. However, researchers assigned participants to flight training sections (i.e., flight blocks) that delivered only the prescribed curriculum to their respective group. ERAU's flight training focus is to produce pilots well prepared for employment as professional pilots. Consequently, specific skills (e.g., cockpit flows and call outs) are standardized and embedded into all flight-training curricula.

The experimental and control curricula were subdivided into units. The participants completed a unit of their group's curriculum during an assigned flight block. Each task had a prescribed training standard graduated in nature to progress from a lower initial level of ability to a higher-level ability as prescribed by PTS. For example, a participant in an early unit would not be expected to land within PTS standards in order to receive a passing grade for the unit. In all cases, participants in the experimental group must perform to PTS standards for a task

before attempting that task in an airplane. Units were arranged in such a way that participants were sequenced from FTD-based training to airplane-based training in order to ensure achievement of PTS standards in simulation prior to airplane flight. The sequencing of the curriculum allowed pilots to proceed to subsequent units, but they were not allowed to attempt tasks that had not yet been performed to PTS standards in the FTD. Each unit was graded and iterations were recorded by the participant's assigned instructor pilot. After the unit was complete, the instructor pilot deposited the iteration slip in a designated location for processing by the researchers.

Results

Researchers calculated mean group iterations required to perform to PTS standards for each task (see Table 1). The data were corrected through a logarithmic transformation (see Table 2) to address the restriction of range issue in the data in which there cannot be fewer than one trial to task completion in actual flight. Skew and kurtosis (kurt) are indices of normality in the data. Skew and kurtosis values higher than two for either indicate non-normal data (Aron & Aron, 1999). For a number of the variables under investigation in this study values for skew and kurtosis indicated non-normality. Logarithmic transformations are a typical method used to correct the data toward a normalized data distribution that is an assumption necessary for MANOVA analyses. The skew and kurtosis values reduced from an average of 1.16 and 1.54 respectively for the untransformed data to 0.22 and -0.23 for the transformed data (see Tables 1 and 2). MANOVA analyses were performed on the transformed data. Researchers calculated TER values with these mean scores. The researchers accounted for any voids in task iteration data. Missing data in the study were replaced by the respective group mean of iterations for that specific flight task. Approximately 9.75% of the data points were filled in this manner. The normalized data occurred primarily in four tasks: Lost Procedures, Diversion, Rectangular Course, and Soft-field Approach and Landing.

Table 1
Iterations to PTS in the Airplane for the Control and Experimental Groups by Task

	Control				Experimental				Total			
	<u>M</u>	<u>SD</u>	<u>Skew</u>	<u>Kurt</u>	<u>M</u>	<u>SD</u>	<u>Skew</u>	<u>Kurt</u>	<u>M</u>	<u>SD</u>	<u>Skew</u>	<u>Kurt</u>
Preflight Inspection	3.6	1.0	-0.3	-0.7	1.4	0.6	1.2	0.8	2.4	1.4	0.4	0.1
Cockpit Management	3.6	1.3	-0.7	-0.5	1.4	0.8	2.3	4.9	2.4	1.5	0.8	2.2
Engine Starting	4.8	2.3	1.0	0.7	1.5	0.8	2.0	4.1	3.1	2.4	1.5	2.4
Taxiing	6.5	4.9	1.8	4.2	2.1	1.3	1.9	3.6	4.2	4.1	1.8	3.9
Before Takeoff Check	5.9	2.9	0.9	0.7	1.6	0.9	1.4	1.5	3.6	3.0	1.2	1.1
Traffic Patterns	11.7	7.9	0.8	-0.7	4.0	3.2	0.8	-0.3	7.6	7.0	0.8	-0.5

table continued on next page

Normal and Crosswind Takeoff and Climb	6.4	3.7	0.5	-0.2	2.5	2.1	1.6	1.7	4.4	3.5	1.1	0.8
Normal and Crosswind Approach and Landing	20.8	15.4	0.7	-0.4	5.1	3.6	0.5	-0.8	12.5	13.4	0.6	-0.6
Soft-field Takeoff and Climb	2.5	1.4	1.0	0.7	2.3	1.1	0.9	0.9	2.4	1.3	1.0	0.8
Soft-field Approach and Landing	3.7	2.0	0.4	-0.9	2.8	1.2	0.0	-0.4	3.2	1.7	0.2	-0.6
Short-field Takeoff and Max Performance Climb	2.4	1.4	0.8	0.8	1.9	0.8	1.2	2.9	2.1	1.1	1.0	1.9
Short-field Approach and Landing	4.0	2.8	1.6	1.7	2.8	0.9	-0.7	-0.5	3.4	2.1	0.4	0.6
Forward Slip to a Landing	2.9	2.1	0.7	-0.7	1.6	0.7	0.9	-0.2	2.2	1.6	0.8	-0.5
Go-Around/Rejected Landing	3.2	2.9	1.7	2.6	1.6	1.0	2.4	6.7	2.4	2.2	2.0	4.7
Steep Turns	4.4	3.4	0.7	-0.1	2.6	1.4	0.9	0.3	3.4	2.7	0.8	0.1
Rectangular Course	2.6	1.3	1.1	1.6	2.0	0.7	-0.3	-1.0	2.3	1.1	0.4	0.3
S-Turns	4.1	2.8	0.6	-1.0	2.5	1.0	-0.3	-1.3	3.3	2.2	0.2	-1.2
Turns Around a Point	3.3	2.3	0.5	-0.4	2.7	1.5	2.5	9.7	3.0	1.9	1.5	4.6
Pilotage and Dead Reckoning	1.5	0.8	2.7	8.2	1.4	0.5	1.9	4.2	1.5	0.7	2.3	6.2
Diversion	1.5	0.5	1.9	3.8	1.5	0.9	1.6	2.5	1.5	0.7	1.7	3.1
Lost Procedures	1.4	0.4	0.7	-1.2	1.2	0.3	1.7	2.8	1.3	0.4	1.2	0.8
Navigation Systems and Radar Services	1.9	1.7	1.3	1.0	1.7	0.7	0.9	-0.1	1.8	1.2	1.1	0.4
Emergency Approach and Landing	4.1	3.6	1.4	2.6	2.2	1.3	1.1	0.9	3.1	2.8	1.3	1.7
Systems and Equipment Malfunctions	2.9	2.3	2.1	5.4	1.8	0.8	0.5	-1.0	2.3	1.8	1.3	2.2
Straight-and-Level Flight (IFR)	2.4	2.0	1.7	2.3	2.1	1.7	0.8	-0.4	2.2	1.8	1.2	1.0
Constant Airspeed Climbs (IFR)	2.4	2.1	2.2	5.7	2.1	1.5	1.8	2.9	2.2	1.8	2.0	4.3

table continued on next page

Constant Air-speed Descents (IFR)	2.5	1.9	1.8	3.2	2.4	2.0	1.8	2.6	2.4	1.9	1.8	2.9
Turns to Headings (IFR)	2.7	1.9	1.3	1.9	2.0	1.8	1.9	3.4	2.3	1.9	1.6	2.6
Recovery from Unusual Attitudes (IFR)	2.0	1.1	0.9	-0.7	1.7	1.2	1.7	2.4	1.8	1.1	1.3	0.9
Radio Communication Navigation Systems/Facilities & Radar Services	5.7	4.6	1.0	0.2	2.8	2.4	1.9	3.8	4.1	3.9	1.5	2.0
Maneuvering During Slow Flight	5.7	4.0	0.8	0.0	2.6	2.0	1.0	-0.2	4.1	3.4	0.9	-0.1
Power-Off Stall	5.4	4.0	0.6	-0.8	2.8	2.9	2.2	4.7	4.0	3.7	1.4	2.0
Power-On Stall	6.4	4.5	0.5	-0.9	3.1	3.2	1.6	1.8	4.7	4.2	1.0	0.4
After Landing, Parking and Securing	3.8	2.1	0.4	-0.7	1.4	0.8	2.3	4.9	2.6	2.0	1.4	2.1

n=18 for control group and 20 for experimental group. N=38 total.

Table 2

Transformed Mean Iterations to PTS in the Airplane for the Control and Experimental Group by Task

	In Control				In Experimental				In Total			
	M	SD	Skew	Kurt	M	SD	Skew	Kurt	M	SD	Skew	Kurt
Preflight Inspection	1.2	0.3	-0.8	-0.3	0.3	0.4	0.9	-0.9	0.7	0.6	0.0	-0.6
Cockpit Management	1.2	0.5	-1.4	1.3	0.2	0.4	1.7	1.8	0.7	0.7	0.1	1.6
Engine Starting	1.5	0.5	-0.2	-0.2	0.3	0.4	1.4	0.9	0.8	0.8	0.6	0.4
Taxiing	1.5	0.7	-0.6	0.8	0.6	0.5	0.6	0.0	1.0	0.8	0.0	0.4
Before Takeoff Check	1.6	0.4	-0.2	0.0	0.4	0.5	0.9	-0.6	0.9	0.8	0.4	-0.3
Traffic Patterns	2.1	0.6	-0.2	-0.2	1.0	0.9	0.1	-1.7	1.5	0.9	-0.1	-0.9
Normal and Crosswind Takeoff and Climb	1.7	0.7	-1.1	1.0	0.6	0.7	0.7	-0.7	1.1	0.9	-0.2	0.2
Normal and Crosswind Approach and Landing	2.7	0.7	-0.4	-0.6	1.3	0.8	-0.4	-1.2	2.0	1.0	-0.4	-0.9
Soft-field Takeoff and Climb	0.8	0.6	-0.1	-0.9	0.7	0.5	-0.2	-0.6	0.7	0.5	-0.1	-0.7
Soft-field Approach and Landing	1.1	0.6	-0.6	-0.4	0.9	0.5	-0.9	-0.4	1.0	0.6	-0.7	-0.4
Short-field Takeoff and Max Performance Climb	0.6	0.4	-0.3	-0.3	0.5	0.4	0.0	-0.3	0.6	0.4	-0.1	-0.3

table continued on next page

Short-field Approach and Landing	1.2	0.7	0.2	0.1	1.0	0.4	-1.3	1.2	1.1	0.5	-0.6	0.6
Forward Slip to a Landing	0.7	0.6	0.1	-1.2	0.4	0.4	0.5	-1.4	0.5	0.5	0.3	-1.3
Go-Around/Rejected Landing	0.8	0.7	0.5	-0.7	0.3	0.5	1.2	0.9	0.5	0.6	0.9	0.1
Steep Turns	1.2	0.5	-0.7	0.9	0.8	0.6	0.0	-1.0	1.0	0.6	-0.4	-0.1
Rectangular Course	0.8	0.5	-0.2	-0.1	0.6	0.4	-0.6	-1.2	0.7	0.5	-0.4	-0.7
S-Turns	1.2	0.8	-0.2	-1.1	0.8	0.5	-0.7	-1.1	1.0	0.7	-0.5	-1.1
Turns Around a Point	0.9	0.5	-0.5	-0.3	0.8	0.5	-0.1	1.1	0.9	0.5	-0.3	0.4
Pilotage and Dead Reckoning	0.2	0.3	1.8	3.9	0.3	0.3	1.2	1.1	0.2	0.3	1.5	2.5
Diversion	0.3	0.3	1.2	1.0	0.2	0.2	1.1	0.9	0.2	0.3	1.1	1.0
Lost Procedures	0.3	0.3	0.5	-1.3	0.2	0.2	1.2	1.2	0.2	0.3	0.9	-0.1
Navigation Systems and Radar Services	0.3	0.4	0.9	-0.6	0.4	0.4	0.4	-1.0	0.4	0.4	0.6	-0.8
Emergency Approach and Landing	1.0	0.6	-0.2	0.0	0.6	0.6	0.2	-1.2	0.8	0.6	0.0	-0.6
Systems and Equipment Malfunctions	0.7	0.6	0.5	0.2	0.5	0.4	0.1	-1.7	0.6	0.5	0.3	-0.8
Straight-and-Level Flight (IFR)	0.6	0.7	0.8	-0.7	0.5	0.5	0.4	-1.5	0.5	0.6	0.6	-1.1
Constant Airspeed Climbs (IFR)	0.6	0.7	0.8	-0.1	0.6	0.6	0.7	-0.4	0.6	0.6	0.8	-0.3
Constant Airspeed Descents (IFR)	0.7	0.7	0.5	-0.6	0.6	0.7	0.9	-0.3	0.6	0.7	0.7	-0.5
Turns to Headings (IFR)	0.7	0.7	0.3	-1.3	0.3	0.5	1.3	0.4	0.5	0.6	0.8	-0.4
Recovery from Unusual Attitudes (IFR)	0.5	0.6	0.4	-1.4	0.4	0.5	1.1	-0.1	0.5	0.5	0.7	-0.8
Radio Communication Navigation Systems/Facilities & Radar Services	1.4	0.9	-0.3	-1.0	0.7	0.7	0.6	-0.7	1.0	0.9	0.2	-0.9
Maneuvering During Slow Flight	1.4	0.7	-0.4	-0.5	0.7	0.7	0.4	-1.6	1.0	0.8	0.0	-1.0
Power-Off Stall	1.3	0.8	-0.5	-0.6	0.6	0.8	1.0	-0.3	1.0	0.9	0.2	-0.4
Power-On Stall After Landing, Parking and Securing	1.2	0.7	-0.6	-0.5	0.2	0.4	1.7	1.8	0.7	0.7	0.5	0.7

n=18 for control group and 20 for experimental group. N=38 total.

MANOVA

Researchers calculated a MANOVA to determine if the number of flight iterations performed in airplane flight to achieve PTS was significantly lower for the

experimental group. Tasks with significantly lower mean iterations for the experimental group are noted with an asterisk in Table 3. There were no tasks with significantly higher mean iterations for the experimental group in the airplane. For all dependant variables $p=0.05$ with 1, 36 degrees of freedom. A MANOVA analysis was selected for these data to reduce the possibility of a Type I error given the large number of dependant variables.

Table 3
Transfer Effectiveness Ratio (TER) from FTD Flight to Airplane Flight and MANOVA results by PTS Task

	TER	F	p
Preflight Inspection*	0.64	76.98	0.00
Cockpit Management*	0.72	37.84	0.00
Engine Starting*	0.59	67.16	0.00
Taxiing*	0.77	19.58	0.00
Before Takeoff Check*	0.82	71.75	0.00
Traffic Patterns*	2.19	17.58	0.00
Normal and Crosswind Takeoff and Climb*	0.57	18.40	0.00
Normal and Crosswind Approach and Landing*	2.1	31.76	0.00
Soft-field Takeoff and Climb	0.06	0.10	0.76
Soft-field Approach and Landing	0.32	1.45	0.24
Short-field Takeoff and Max Performance Climb	0.13	0.63	0.43
Short-field Approach and Landing	0.27	1.17	0.29
Forward Slip to a Landing*	0.48	5.67	0.02
Go-Around/Rejected Landing*	0.51	4.23	0.05
Steep Turns*	0.32	4.22	0.05
Rectangular Course	0.32	2.77	0.10
S-Turns	0.53	3.30	0.08
Turns around a Point	0.2	0.20	0.66
Pilotage and Dead Reckoning	0.09	0.10	0.75
Diversion	-0.02	1.06	0.31
Lost Procedures	0.18	1.27	0.27
Navigation Systems and Radar Services	0.1	0.63	0.43
Emergency Approach and Landing*	0.69	4.97	0.03
Systems and Equipment Malfunctions	0.41	2.57	0.12
Straight-and-Level Flight (IFR)	0.09	0.45	0.51
Constant Airspeed Climbs (IFR)	0.1	0.09	0.77
Constant Airspeed Descents (IFR)	0.05	0.13	0.72

table continued on next page

Turns to Headings (IFR)*	0.3	3.99	0.05
Recovery from Unusual Attitudes (IFR)	0.09	0.72	0.40
Radio Communication Navigation Systems/Facilities & Radar Services*	0.82	5.50	0.02
Maneuvering During Slow Flight*	0.38	10.75	0.00
Power-Off Stall*	0.27	6.82	0.01
Power-On Stall*	0.34	9.79	0.00
After Landing, Parking and Securing*	0.74	26.92	0.00

* indicates significant F value.

Discussion

The transfer effective ratios (TERs) (see Table 3) indicated that 33 out of 34 tasks had positive transfer from FTD flight to aircraft flight. For 18 of these 34 tasks, the experimental group required significantly fewer iterations to achieve PTS standards in the airplane after they trained to standard in the FTD. The positive direction of the TERs, coupled with significantly lower number of iterations to achieve PTS in the airplane by the experimental group, strongly suggests that these FTDs are an effective means for training ab-initio student pilots. Some tasks were more effectively trained than others were. Flight training developers need to weigh the issue of effective transfer when determining the curricula selected to meet training needs. At times, tasks with apparently low levels of transfer effectiveness are most effectively trained in simulation when safety and/or monetary savings are considered. There are several possible explanations addressing whether a task can be effectively trained in a simulation device with the functionality of these FTDs for ab-initio flight students. Based on direct observations and instructor pilot interviews, the researchers categorized factors that indicate potential explanations for the degree of transfer in multiple tasks.

Visual Fidelity

The FTDs served primarily for training visual flight rule (VFR) tasks. Pilot perceptions of vection (i.e., a visually induced false sensation of self-movement) occur primarily by sensing movement of objects in the peripheral vision. The motion parallax effect afforded by the enhanced visual scene in the simulation enhances perceptions of vection. Tasks performed in close proximity to detailed and well-developed 3-dimensional (3-D) graphic artwork in the virtual world typically indicated higher levels of transfer when compared to those practiced in less developed areas of the virtual world. These tasks included Taxing, Traffic Patterns, Normal and Crosswind Take-Off and Climb, Normal and Crosswind Approach and Landing, Forward Slip to Landing, Go Around/Rejected Landing, Emergency Approach and Landing, and After Landing Parking and Securing. Students performed these tasks in the highly developed virtual flight environment immediately surrounding Daytona Beach International Airport (KDAB). The significance of these tasks suggests that the high fidelity visual display, in conjunction with the well-developed 3-D graphic artwork, had a positive effect on transfer from FTD-based flight to airplane-based flight. Researchers found ToT was not significant in the well-developed 3-D graphic virtual environment for four tasks.

These tasks were Soft-field Take Off and Climb, Soft-field Approach and Landing, Short-field Off Take Off and Max Performance Climb, and Short-field Approach and Landing. Researchers hypothesize these four task's inherent degree of difficulty (i.e., the tasks are difficult to master regardless of application of a real or virtual training environment) affected performance and further research is necessary to isolate the causes.

While fidelity of the visual scene improved ToT for multiple tasks, the impact of a high level of visual fidelity appeared to have minimal positive transfer in other tasks. Rectangular Course, S-Turns, and Turns Around a Point are ground reference maneuvers performed in the practice areas to the northwest and southeast of KDAB between 600 and 1,000 feet above ground level (AGL). Developers optimized the FTD visual system for flight at 3,000 feet AGL and above. This is due to the nature of the satellite imagery underlying the virtual world. As pilots descend to lower altitudes, and in the absence of 3-D graphics, visual fidelity is compromised. This impairment to visual fidelity can account for the minimal positive transfer for these tasks.

Pilotage and Dead Reckoning, Diversion, and Lost Procedures were not significantly different between the two groups for transferring skills to airplane flight. The researchers hypothesize the optimization of the visual system at 3000 feet AGL in conjunction with a lack of detailed 3-D graphical art work across a relatively long (i.e., 150 nautical mile) flight route may have failed to deliver the cues necessary to effectively train these tasks that are heavily reliant on external visual cues. Diversion was the only task indicating negative transfer. This negative transfer was not significant for this task, but warrants further investigation. Diversion might be a difficult task in the virtual environment due to its inherent lack of well-developed visual virtual landmarks. Being diverted in the real world environment (i.e., flying into new airspace) is greatly aided by a surplus of visual landmarks that might be useful during navigation.

Procedural Similarity

The theory of Identical Elements as initially stated by Thorndike (1906) suggested that transfer only occurs in the presence of specific common elements. High fidelity in the forms of physical fidelity (e.g., the FTD's real Cessna C-172S cockpit), cognitive fidelity (e.g., instructor pilots role playing air traffic/air traffic control and ab-initio pilot realistically experiencing cognitive work loading during training), control loading fidelity, (e.g., realistic force feed back on flight controls), and aerodynamic fidelity enables the FTD to properly mimic airplane flight. The researchers deem certain PTS tasks are highly procedural in nature and are readily replicated in the simulated flight training environment used at ERAU for research. These tasks include: Preflight Inspection, interior cockpit only; Cockpit Management; Engine Starting; Before Take-Off Check; Radio Communication Navigational Systems/Facilities and Radar Services; Traffic Patterns; Steep Turns; and After Landing Parking and Securing.

Difficulty of Tasks

Several flight tasks necessitate higher levels of skill than others (e.g., it is inherently more difficult to perform a Short-Field Approach and Landing than a Normal and Crosswind Approach and Landing, the PTS standard for a Short-

Field Approach and Landing is 200 feet while the PTS standard allows 400 feet for a Normal and Crosswind Approach and Landing). Typically, ab-initio pilots master these tasks during the later stages of their training. Soft-Field Takeoff and Climb, Soft-Field Approach and Landing, Short-Field Takeoff and Climb, Short-Field Approach and Landing are taught late in the curricula. These tasks proved more difficult to master for participants in both the experimental and control groups. Positive transfer occurred for each task but the difference between the simulation group and the control group was not significant. The data suggested that these tasks are difficult to achieve no matter where they were first learned. Training to standard in the FTD did not seem to mitigate the difficulty of mastering these tasks. The sequencing of training tasks in the curricula had the goal of adhering to the building block principle of learning (i.e., a concept where knowledge and skills are best learned based on previous associated learning experiences) (Federal Aviation Administration, 1999).

Visual Scanning and Response

The tasks of Slow Flight, Power-Off Stalls, and Power-On Stalls also showed a significantly lower number of iterations necessary for the experimental group to achieve PTS standards in the airplane when compared to the control group. While in actual flight performance of these tasks rely heavily upon the students' ability to perceive and respond to proprioceptive stimulation for pitch attitude and the sensation of falling. Participants learning these tasks in the FTD perform in the absence of proprioceptive stimulation. They were forced to rely exclusively upon their visual sense to determine the aircraft state as it approaches the indicated airspeeds (IAS) that result in a stalled condition. The students' ability to maneuver during slow flight and properly recovering from a stalled flight condition was positively affected by their training in the FTD. Enhanced attention to the flight instruments may allow the participants to perform these tasks in flight after training in the FTD. Ab-initio students may learn to scan more efficiently between instruments and the out-of-the-cockpit visual scene while mastering these tasks in the FTD.

Dynamic Flight Environment

The FTDs incorporate a degree of unstable air mass modeling. The weather modeling is optimized to replicate flight conditions experienced during relatively stable departure, enroute, and approach stages of flight. Complex and changing combinations of updrafts, downdrafts, crosswinds, and headwinds tremendously affect control inputs necessary to perform flight tasks and remain within PTS prescriptions. Previous research examining the transfer of skills from simulated flight to real flight under simulated instrument meteorological conditions (IMC) has not addressed performance by ab-initio pilots. This previous research typically addressed performance by pilots with at least Private Pilot certification. More experienced pilots are already familiar with the feel of the aircraft and how it will react during each maneuver. Straight-and-Level Flight, Constant Airspeed Climbs, Constant Airspeed Descents, Turns to Headings, and Recovery from Unusual Attitudes are tasks taught to ab-initio pilots under instrument flight rules (IFR) in simulated IMC. None of the basic instrument tasks showed a significant difference in airplane iterations between the two groups, with the exception of Turns to Headings in the transformed data. The researchers hypothesize that the FTD does not mimic all of the complexities of air currents experienced in actual flight.

Conclusion

The researchers' purpose for this study was to quantify the transfer effectiveness of training in FTDs to performance in an actual airplane. This study used simulated flight as the primary means of training ab-initio pilots as they earned a Private Pilot's certificate. The study was longitudinal in nature. It followed the performance of participant pilots from a novice condition to certification by the FAA as a Private Pilot. This study included elements that differentiate it from other studies in that it included ab-initio pilots culminating in FAA certification as a Private Pilot under a Part 142 approved curriculum. The curriculum was primarily comprised of flight training with simulated flight.

The analysis of the data and direct observations of performance lead the researchers to believe training ab-initio pilots with an FTD that has the functionality and fidelity of the devices in use at ERAU can be effective. Transfer of training was positive in 33 out of 34 tasks and significantly different from the control group in 18 out of 34 tasks. Optimizing flight curricula to capitalize on the strengths and weaknesses of the device is critical to flight training. Embry-Riddle Aeronautical University is in the process of applying the results of this research to its flight curricula developmental process. Now and in the future, ERAU ab-initio pilots will train with a flight curriculum that relies heavily upon FTD flight.

References

- Anderson, R. P., & Macchiarella, N. D. (2004). Nonmotion flight simulation development of slow flight and stall tasks for use in ab-initio flight training. *Proceedings of the Human Performance, Situation Awareness and Automation Technology II (HPSAA II) Conference*. Daytona Beach, FL, II, 180-184.
- Aron, A., & Aron, E.N. (1999). Strategies when population distributions are not normal: Data transformations, rank-order tests, and computer-intensive methods. In *Statistics for psychology* (2nd ed., pp. 458-487). Upper Saddle River, NJ: Prentice Hall.
- Beaubien, M. J., Baker, D. P., & Salvaggio, A. N. (2004). Improving the construct validity of line operational simulation ratings: Lessons learned from the assessment center. *International Journal of Aviation Psychology, 14*(1), 1-17.
- Brown, Y. J., Cardullo, F. M., & Sinacori, J. M. (1989). *Need-based evaluation of simulator force and motion cuing device* (No. 89-3272-CP). Reston, VA: American Institute of Aeronautics and Astronautics.
- Champney, R., Milham, L., Carroll, M. B., Stanney, K. M., Cohn, J. (2006). A method to determine optimal simulator training time: Examining performance improvement across the learning curve. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. San Francisco, CA, 2654-2658
- Embry-Riddle Aeronautical University. (2003). *Cessna 172S standard operating procedures*. Daytona Beach: Embry-Riddle Aeronautical University.
- Embry-Riddle Aeronautical University. (2005a). *Private pilot SEL training manual*. Unpublished Curriculum, Approved by FSDO-15. Embry-Riddle Aeronautical University, Daytona Beach, FL.
- Embry-Riddle Aeronautical University. (2005b). *Private pilot TCO*. Unpublished Curriculum, Approved by FSDO-15. Embry-Riddle Aeronautical University, Daytona Beach, FL.

- Federal Aviation Administration. (1992). *Airplane flight training device qualification*. Advisory Circular (AC) 120-45A, Author.
- Federal Aviation Administration. (1999). *Aviation instructor's handbook*. (FAA-H-8083-9). Author.
- Federal Aviation Administration. (2002). *Private pilot practical test standards for airplane* (FAA-S-8081-14AS). Author.
- Gerathewohl, S. J., Mohler, S. R., & Siegel, P. V. (1969). *Fidelity of simulation and transfer of training: A review of the problem* (No. AM 69-24): Federal Aviation Administration, Office of Aviation Medicine.
- Go, T. H., Bürki-Cohen, J., & Soja, N. N. (2000). *The effect of simulator motion on pilot training and evaluation* (No. AIAA 2000-4296). Denver, CO: AIAA Modeling and Simulations Technologies Conference.
- Hays, R. T., Jacobs, J. W., Prince, C., & Salas, E. (1992). Flight simulator training effectiveness: A meta-analysis. *Military Psychology*, 4(2), 63-74.
- Jacobs, R. S., & Roscoe, S. N. (1975). Simulator cockpit motion and the transfer of initial flight training. *Proceedings of the 19th Meeting of the Human Factors Society*, Santa Monica, CA.
- Patrick, J. (2003). Training. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology*. NJ: Mahwah: Lawrence Erlbaum Associates, Publishers.
- Rantanen, E., & Talleur, D. A. (2005). Incremental transfer and cost effectiveness of ground-based trainers in university aviation programs. *Proceedings of the Human Factors and Ergonomics Society, 49th Annual Meeting*, Orlando, FL, 764-768.
- Roscoe, S. N., & Williges, B. H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), *Aviation psychology* (pp. 182-193). Ames: Iowa State University Press.
- Stewart, J. E., Dohme, J. A., & Nullmeyer, R. T. (2000). U.S. Army initial entry rotary-wing transfer of training research. *The International Journal of Aviation Psychology*, 12(4), 359-375.
- Thorndike, E. L. (1906). *The principles of teaching based on psychology*. New York: A.G. Seiler.
- Waag, W. L. (1981). *Training effectiveness of visual and motion simulation* (No. AFHRL-TR-79-72). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.
- White, A., & Rodchenko, V. (1999). *Motion fidelity criteria based on human perception and performance* (No. AIAA-99-4330). Vienna, VA: American Association of Aeronautics & Astronautics.

The Evaluation of the Decision Making Processes Employed by Cadet Pilots Following a Short Aeronautical Decision-Making Training Program

Wen-Chin Li

*Department of Psychology
National Defense University, Republic of China*

and

Don Harris

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

Abstract

Many aeronautical decision-making (ADM) mnemonic-based methods exist. However, there is no empirical research that suggests that they are actually effective in improving decision-making. Klein (1993), in his study of naturalistic decision making, suggested that the decision-making process centers around two processes: situation assessment to generate a plausible course of action and mental simulation to evaluate that course of action for risk management. In this study a short, ADM training course was constructed around two mnemonic methods, SHOR (Stimuli, Hypotheses, Options, and Response) and DESIDE (Detect, Estimate, Set safety objectives, Identify, Do, Evaluate). Forty-one pilots from the Republic of China Tactical Training Wing participated: half received a short ADM training course and half did not. After training, the procedural knowledge underpinning their Situation Assessment and Risk Management ability, two skills essential for successful decision-making, were evaluated using pencil and paper-based knowledge tests based upon several demanding tactical flight situations. These scenarios were designed to encompass the six basic types of decision making described by Orasanu (1993); go/no go decisions; recognition-primed decisions; response selection decisions; resource management decisions; non-diagnostic procedural decisions, and decisions requiring creative problem-solving. The results show gains attributable to the decision making training course in both situation assessment and risk management skills. The results strongly suggest that ADM is trainable and such a training course is effective in improving the bases of in-flight decision-making.

Introduction

Aeronautical decision-making (ADM) is defined by the FAA (1991) as 'a systematic approach to the mental process used by aircraft pilots to consistently determine the best course of action in response to a given set of circumstances'

Requests for reprints should be sent to Kay Chisholm, FAA Academy, AMA-530, P.O. Box 25082, Oklahoma City, OK 73125. E-mail to kay.chisholm@faa.gov.

(Hunter, 2003). Jensen (1995) defined pilot judgment as 'the mental process that pilots use in making decisions'. Both definitions implicitly include both process and outcome. For military pilots operating in a hostile environment, the normal hazards of aviation are compounded by the enemy's intent for the destruction of the aircraft. Fischer, Orasanu, & Wich (1995) suggested that risk and time pressure are situational variables that further constrain the decision process, as risk and time pressure may call for an immediate response whether or not the problem was fully understood. Minimal risk levels and fewer time constraints, in contrast, permit additional diagnostic actions or the deliberation of options.

Klein (1993), in his study of naturalistic decision making suggested that the decision-making process centers around two processes: situation assessment, which is used as a precursor to generate a plausible course of action and mental simulation to evaluate that course of action for risk management. If a pilot recognizes there is sufficient time for making wide-ranging considerations, s/he will evaluate the dominant response option by conducting a mental simulation to see if it is likely to work. If there is not adequate time, the pilot will tend to implement the course of action that experience (if any) dictates is the most likely to be successful.

Endsley (1997) defines situation awareness (SA) as 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the future'. In the dynamic tactical environment, effective decision-making is highly dependent on situation awareness, which has been identified as a critical decision component (Endsley & Bolstad, 1994). Situation assessment is the process by which the state of situation awareness is achieved and is a fundamental precursor to situation awareness, which is itself the precursor for all aspects of decision-making (Nobel, 1993; Prince & Salas, 1997).

Jensen, Guilke & Tigner (1997) suggested that risk management should be a key part of the decision-making process. Risk assessment feeds into decision making in two ways: during the assessment of the precipitating threats and in evaluating potential courses of action. Janis and Mann (1977) proposed that a good decision-making process is one in which the decision maker successfully accomplishes the collection of information about a wide range of alternatives, carefully assesses the risks and benefits of each course of action, and prepares contingency plans for dealing with known risks.

Tactical flight training has many aspects that challenge the quality and processes of pilots' in-flight decision-making. In addition to the tasks and situations faced by the pilot of a civil aircraft, military pilots must perform a wide range of other tasks in addition to flying their aircraft safely. Their primary task may be to intercept offensive aircraft or to deliver weapons, troops, or equipment. Often the act of flying the aircraft *per se* in a hostile environment becomes a secondary task. As a result, military pilots must learn to make decisions related to mission performance in addition to those related to flying the aircraft *per se* (Kaempff &

Orasanu, 1997). Flying advanced fighter aircraft has made increasing demands on pilots' cognitive abilities as the complexity of cockpit systems and the tactical situation has grown. There is now a requirement for decision-making training to be incorporated into tactical training programs (Li, Harris & Yu, 2005a). Furthermore, many accidents are either wholly or partially attributable to poor decision-making (Li, Harris & Yu, 2005b). However, at the present time, there is little or no formal training available for military pilots in the ROC Air Force or elsewhere offering heuristics, procedures, or advice about making effective decisions under high pressure and in a time-limited, tactical situation.

Many researchers have suggested that ADM is trainable (Endsley, 1993; Klein, 1993 & 1997; Orasanu, 1993; Prince & Salas, 1997; Li & Harris, 2005). Buch and Diehl (1984) found that judgment training produced significantly better decisions among civil aviation pilots. Connolly, Blackwell & Lester (1989) observed that decision-making skills could be improved by the use of judgment training materials coupled with simulator practice. However, Orasanu (1993) suggested that generic training techniques to improve all-purpose decision-making skills would not be successful. She suggested that different component skills were involved when making six different basic types of decisions (go/no go decisions; recognition-primed decisions; response selection decisions; resource management decisions; non-diagnostic procedural decisions; and creative problem-solving).

There are a number of strategies embodied in mnemonics or acronyms describing the processes and procedures concerned with ADM. These have been developed in recent years by researchers and used by pilots to support ADM 'best practice' (e.g. Wohl, 1981; Maher, 1989; Klein & Woods, 1993; Hormann, 1995; Oldaker, 1996; Jensen, 1997; David, 1997; Murray, 1997; Orasanu, 1997; O'Hare, 2003). The common aim of these techniques is to encourage a systematic approach to decision-making that should be less affected by the human nature and should also reduce the cognitive work for pilots (O'Hare, 2003). However, there is a lack of hard empirical research demonstrating the effectiveness of these ADM mnemonic methods.

Li & Harris (2005) undertook a study to identify the best ADM mnemonic-based methods for training military pilot's decision-making. From the results of this study it was found that SHOR (Wohl, 1981) was rated as being the best ADM mnemonic in time-limited and critical, urgent situations. DESIDE (Murray, 1997) was regarded as superior for knowledge-based decisions which required more comprehensive considerations but also had more time available to do so. The SHOR mnemonic (Wohl, 1981) consists of four steps: Stimuli, Hypotheses, Options, and Response. It was originally developed for use by U.S. Air Force tactical command and control, where decisions were required under high pressure and severe time constraint. In this situation, decisions require near-real-time reactions involving threat warning, rescheduling, and other types of dynamic modification. The SHOR methodology is basically an extension of the stimulus-response paradigm of classical behavioral psychology developed to deal with two

aspects of uncertainty in the decision-making process, information input uncertainty (which requires hypothesis generation and evaluation) followed by the evaluation of the consequences of actions, which creates the requirement for option generation and evaluation. DESIDE (Murray, 1997) was developed on a sample of South African pilots and comprises six steps, Detect, Estimate, Set safety objectives, Identify, Do, Evaluate. The DESIDE method is a practical application to aid pilots in making in-flight decisions adapted from the conflict-theory model of Janis and Mann (1977).

O'Connor, Flin, Fletcher & Hemsley (2002) described several methods for the evaluation of CRM (Crew Resource Management) and ADM training, including the use of simulator/LOFT checks; self/peer/360 degree appraisals; the assessment of technical performance; the analysis of confidential reports and the use of knowledge assessment tests. The standard method for the assessment of the knowledge-based elements is normally a pencil and paper based test. This provides a reasonably quick and simple way of evaluating knowledge acquisition. The following study evaluates the effectiveness of a short ADM training course delivered to ROC Air Force cadet pilots based around the SHOR and DESIDE ADM mnemonic-based methods using a pencil-and-paper knowledge based approach. The ADM training course (described in more detail in the following section) also provided advice concerning which ADM approach was most suitable in any given situation. It is argued that the decision making training program delivered requires assessment in two aspects: the actual decision-making *performance* of students on completion of the training and an assessment of the *process* by which they arrive at their decision. In this paper, emphasis is placed on the evaluation of the pilots' decision-making *process* and the quality of the decision based around the dimensions of situation assessment and risk management. The results of the *product*-based measures of the training program, evaluated using decision scenarios re-created in a full-flight simulator are reported elsewhere (Li, Harris & Yu, 2005b). While these simulator trials could assess the products of the ADM training program in a time-pressured, real-time environment they had severe limitations in establishing if the processes taught within the training course were being applied appropriately, hence the requirement for the knowledge-based pencil and paper tests.

Furthermore, when evaluating decision-making efficacy, Baron and Hershey (1988) suggested that the study of 'outcomes' shows a tendency of people to assess the correctness of their decision-making with regard to the outcome of the decision. However, good decisions can lead to bad outcomes (and vice versa) especially when operating in a probabilistic environment, such as aviation. Decision makers cannot infallibly be graded by their results (Brown, Kahr, & Peterson, 1974). A good decision cannot guarantee a good outcome. All in-flight decisions are made under uncertainty. Evaluating a decision as good (or not) must depend as much on the stakes and the processes employed, not just simply on the outcome. Hence, in this study the evaluation of the effectiveness of decision-making training is based around the decision-making adjuncts of situation assessment and risk management measures rather than simply on assessing the outcomes of the decisions made.

Using Kirkpatrick's (1976, 1998) hierarchy for training evaluation, the current study assesses the product of the training interventions at the second level of evaluation (learning). The pencil-and-paper based evaluation of the ADM training program delivered is specifically concerned with establishing if the participants have acquired the decision-making procedural knowledge as a result of attending the training course. It was hypothesized that the provision of ADM training would produce superior situation assessment and risk management performance (two key factors underpinning effective decision making) in a range of in-flight decision-making scenarios encompassing Orasanu's (1993) six decision-making categories.

Method

Participants

Forty-One male participants from ROC Air Force Tactical Training Wings participated in the study. The flying experience of participants was between 220 and 354 hours with an average of 292 hours. Participants were randomly divided into two groups, 21 pilots in the experimental (trained) group and 20 pilots in the control (untrained) group.

The Contents of ADM Training Programs

The results from a previous study by Li and Harris (2005) found that just two mnemonic-based methods provided a suitable basis for all aspects of ADM training. These methods encompassed all the requirements of the six basic decision making situations. SHOR (Wohl, 1981) was regarded as being the best for time-limited and urgent situations; DESIDE (Murray, 1997) was regarded as being superior for guiding knowledge-based decisions needing more comprehensive consideration. These two mnemonic methods formed the basis of the ADM training programs. The objective of the training course was to equip trainees with the procedural knowledge required to use these methods.

The training program commenced with an introduction to ADM theories, including the Recognition-Primed Decision Model of Rapid Decision Making (Klein, 1993); The ARTFUL Decision Maker: A Framework Model for Aeronautical Decision Making (O'Hare, 1992); Conflict-theory Decision Making Model (Janis & Mann, 1977); a Model of Situation Awareness in Dynamic Decision-making (Endsley, 1997); and the Decision Process Model (Orasanu, 1995). This was followed by a description of the content and method of application of the SHOR and DESIDE ADM mnemonic-based methods. To optimize decision making training effectiveness it was also necessary to instruct pilots with regard to which technique was the most appropriate to apply in any given circumstance.

Following this, participants underwent a period of supervised practice in the classroom in the application of SHOR and DESIDE in flight situations exemplifying the six basic types of decision making scenario described by Orasanu (1993). Finally, the application of ADM in military aviation was described and the participants who participated in the training course were de-briefed. The ADM training program lasted approximately four hours in total.

Scenarios for the Assessment of ADM Training Effectiveness

To develop scenarios for assessing the effectiveness of the ADM training intervention, which corresponded to Orasanu's (1993) six decision making categories, six focus groups were conducted, one for each scenario. Each focus group comprised one human factors specialist and three senior instructor pilots. The purpose of these focus groups was to verify that the scenarios used in the pre-training and post-training evaluation of decision-making (which were developed from the ROCAF accidents and incidents database) corresponded to the appropriate categories of decision-making and were of equivalent difficulty. Further details of the process validating the selection of the scenarios for each generic decision type can be found in Li & Harris (2005).

To negate practice effects, different (but equivalent) scenarios were used in the evaluations pre- and post ADM training. These focus groups also ensured enough detail was available for pilots to be able to make a decision and hence to evaluate their decision-making performance. These scenarios developed were as follows.

Go/no go decision-making scenario. Go/no go decisions are made under severe time pressure and involve considerable risk; the amount of thinking should be minimal. Orasanu (1993) suggests that training design should focus on developing perceptual patterns in memory that constitute the conditions for the required action. However, they should be trained under realistic time pressure and the training scenarios should include additional contingencies that require more complex risk assessment.

Pre-training scenario: F-5E No. 2 wingman has to make a decision as the No. 1 (Leader) abandons a tactical formation take-off at 145 knots.

Post-training scenario: F-5E No. 2 wingman practicing tactical formation training; during the take off run with the throttles increased to maximum, No.1 (leader) suddenly slants seriously towards the No.2.

In both the above scenarios, the pilots had to make a decision under time pressure with high risk. The patterns of events needed to be recognized and pre-set responses needed to be executed swiftly. The cognitive activities required of the pilots were essentially perceptual and interpretive.

Recognition-primed decision-making scenario. Recognition-primed decisions are described by Orasanu (1993) as the recognition of the situational patterns that serve as inputs to condition-action rules, but which also require the decision maker to learn the response side of the rule and its link to that condition.

Pre-training scenario: F-5E right engine fails as a result of Foreign Object Damage just as the nose gear leaves the ground at a speed of 165 knots.

Post-training scenario: F-5E solo, after taking off at 500 feet, pilot hears two unusual sounds from the engines and feels the aircraft shake. Engine exhaust gas temperature is increased, and RPM decreased.

As noted earlier in the Introduction, Klein (1993) suggested that recognition-primed decision focuses on the two processes of situation assessment and mental simulation. If there is no time to make a considered response (as in the case of both the above scenarios) the pilot will implement the rule that experience has determined will be the most likely to be successful. These situations require more conscious cognitive processing than go/no go decisions (cf. Reason's rule-based errors; Reason 1990).

Response selection decision-making scenario. Response selection decisions involve a single option that must be selected from a set of possible options; pilots must identify the possible options and evaluate them in terms of how well they satisfy the goals and meet constraints. Often they must consider trade-offs among competing goals, which are satisfied by different options.

Pre-training scenario: No. 4 wingman in a tactical formation of F-5Es is required to make a decision when No. 1 (Leader) becomes lost in clouds during formation flight (three feet distance between wing tips of the four fighters).

Post-training scenario: F-5E leader was maintaining loose formation with No. 2 on the left, at 13,000 feet; the Ground Intercept Controller reports an unidentified aircraft at one o'clock and 5 miles away. At the same time No.2 makes visual contact with an airliner in front and head-on at 3 miles away with same altitude and approaching fast (leader had no orders).

In both scenarios, the wingman has to make a decision to choose a response to deal with an impending hazard. Although these are not urgent situations, pilots may perceive the potential risk in front of them to be very high and choose an inappropriate course of action. However, once the nature of the potential threat is identified there are detailed procedures available from their training of how to deal with the situation.

Resource management decision-making scenario. Resource management decisions involve the *relative* priorities of various tasks, especially critical ones. Skills relevant to this type of decision include estimation of the time required to complete the various tasks, knowledge of the interdependencies among tasks, and scheduling strategies.

Pre-training scenario: F-5E leader of four aircraft needs to make a decision for the No.3 and No. 4 aircraft when a 'no joy' call (no visual contact with No. 1 and No. 2) is made and No. 2 calls 'one opposing target approaching on 12:30 o'clock with same altitude'. This occurs during practice of a 2 versus 2 Air Combat Maneuver engagement.

Post-training scenario: Leader and No.2 are practicing basic fighting maneuvers for a gunshot attack; the distance between No. 2 and the leader is only 500 feet, the angle off is over 90 degrees. The possibility of a mid-air collision is high; both aircraft are at 480 knots and same altitude.

Perhaps the most critical issues for resource management decisions are setting the priorities of the responses required to make and implement a decision. In the scenarios described above, the resource allocation problem changes from one of practicing basic fighting maneuvers to one of avoiding a collision. There are certain actions that must be completed within a few seconds to avoid a mid-air crash and they must be prioritized and undertaken in a certain order, such as calling out to alert other traffic prior to climbing, descending, or changing direction.

Non-diagnostic procedural decision-making scenario. Non-diagnostic procedural decisions involve a number of cues falling into a category with no prescribed response. The nature of the problem is unclear and many different types of ambiguous cues may also signal potentially dangerous conditions. Orasanu (1993) suggests that training for this type of decision should involve mainly situation assessment and risk assessment. Cues that signal possible emergencies need to be distinguished from those that are troublesome but not severe enough to precipitate an emergency landing.

Pre-training scenario: Both the leader and wingman in a formation of F-5Es are unable to land at home-base in a 'bingo' (low fuel) situation during instrument flight in bad weather.

Post-training scenario: When an F-5E is finishing Basic Fighting Maneuver training, the Ground Intercept Controller reports that home base weather is worsening. Surplus fuel is down to only 1,400 lb. The pilot asks for weather conditions at alternative airports.

In both the pre-test and post-test scenarios pilots had to evaluate the strengths and weakness of using alternative airfields in deteriorating weather in a 'bingo fuel' situation. There was no clearly defined 'correct' answer. Although the nature of the immediate problem is clear (deteriorating weather at home base), the problems imposed by diverting to an alternate airfield are unclear and deviations from the optimal solution may be required due to the low fuel state.

Creative problem-solving decision-making scenario. Creative problem-solving decisions are the most complex, as they involve both diagnoses to determine the nature of the situation and response generation. Pilots must determine what their goals are, develop a plan and candidate strategies, and evaluate these strategies and actions based on projections of likely outcomes (Orasanu, 1993).

Pre-training scenario: When flying an F-5F both left and right generators warning lights become active during a tactical maneuver.

Post-training scenario: When lowering the landing gear while on the downwind leg the landing gear shaft warning light illuminates, indicating the nose landing gear is abnormal.

In both the decision-making scenarios presented, once the true nature of the problem has been determined (from the indications in the cockpit the pilot was only initially aware of the symptoms of the problem in both cases, not their ultimate cause) they would determine that there were no recommendations in the SOPs/manuals for its resolution, hence a novel solution had to be developed to address the situation.

Procedure

Both experimental (trained) and control (untrained) groups undertook an initial set of pencil and paper based evaluations where they were required to describe how they would deal with each of the problems described in above pre-training decision making scenarios. These evaluations were simply in the form of narrative-based reports describing the steps that they would take when assessing their options and coming to a decision. After these initial tests, the experimental group attended a four-hour 'ADM training Program for military pilots'. The Control group had no such training. Both groups then participated in a further set of pencil and paper evaluations.

To eliminate order effects, the six decision making scenarios were presented in a randomized order in both the pre- and post-training trials. The narrative responses describing the process by which the participants would arrive at their decision were evaluated by a flight instructor with regard to their situation assessment and risk management performance. These dimensions were derived from the earlier study (Li and Harris, 2005) used to select the most appropriate ADM training mnemonic methods. Each aspect of performance was rated using a nine-point Likert-type scale (with a high score of 9 and a low score of 1).

To enhance the reliability of the measures, the same instructor evaluated trainee performance on all occasions. The instructor was trained by an aviation human factors specialist to evaluate performance in the required manner. The narratives describing the decision making process were anonymized before being passed to the flight instructor, thus he was blind to the experimental condition. Furthermore, the instructor took no part in delivering any aspect of the aeronautical decision making training course.

For the evaluation of both Situation Assessment and Risk Management performance in the narrative answers produced, a list of key performance factors (taken from the training manuals) was derived for each scenario. The steps that should be undertaken and sources of information that should be interrogated in each circumstance were listed, these being factors underlying Situation Assessment performance in particular. Emphasis on the risk management dimension was placed upon the generation and analysis of options and the quality of reasoning underlying the pilot's final decision based specifically on the control of risk.

Ethical Approval

This research program was approved by the Ethics committee of Cranfield University. This committee operates to the principles prescribed by the British Psychological Society (the UK professional body for psychologists). Participants

were volunteers and informed of the purpose of the study prior to participating. All data were collected anonymously.

Results

Data

The ADM decision making process that each participant employed was evaluated in all six scenarios in both pre- and post ADM training. In total 492 narrative responses were collected, 246 prior to ADM training taking place and the same number after the training course had been delivered. Two hundred and fifty-two trials were undertaken by the experimental group and 240 by the control group. To re-iterate, the ADM processes described in the narratives produced by the cadet pilots were rated on the dimensions of situation assessment and risk management.

Go/no go Decisions. Irrespective of experimental group, there was no overall difference in situation assessment performance between the pre- and post-test ($F_{1,39}=1.214$; $p=0.277$). There was an effect approaching significance between the trained and untrained group ($F_{1,39}=3.277$; $p=0.078$). The group that had received ADM training tended to outperform the group that had not received training (table 1). The interaction term between the trained/untrained group and pre- post-training was significant ($F_{1,39}=4.355$; $p=0.043$). The group that had received ADM training showed significantly greater gains in the second trial compared to the untrained group. Overall, there was no difference on risk management performance between the pre- and post-test ($F_{1,39}=0.448$; $p=0.507$). There was also no significant difference between the trained and untrained group ($F_{1,39}=2.207$; $p=0.145$). However, there was an effect verging on significance with regard to the interaction term between the trained/untrained group and pre- post-training trial ($F_{1,39}=3.266$; $p=0.078$). The group that had received ADM training showed somewhat greater gains in risk management performance during the second trial compared to the untrained group.

Table 1

Means and Standard Deviations in performance scores in the Go/no go decision-making scenario, broken down by both main effects (pre-test/post-test: trained/untrained) on the measures of situation awareness and risk management

Go/no go decisions		Group	N	Mean	Standard deviation
Situation assessment	Pre-test	Trained	21	5.38	1.20
		Untrained	20	5.25	1.74
	Post-test	Trained	21	6.19	0.98
		Untrained	20	5.00	1.65
Risk management	Pre-test	Trained	21	5.57	1.08
		Untrained	20	5.30	1.53
	Post-test	Trained	21	5.95	1.07
		Untrained	20	5.05	1.23

Recognition-Primed Decisions. There was no difference in situation assessment performance between the pre- and post-test ($F_{1,39}=0.927$; $p=0.342$). There was also no significant difference between the trained and untrained group ($F_{1,39}=1.337$; $p=0.225$). However, there was a significant interaction effect between the trained/untrained group and pre- and post- ADM training trial ($F_{1,39}=9.555$; $p=0.004$). The group that had received ADM training showed significantly greater gains in performance in the second trial compared to the untrained group (table 2). There was no significant difference in risk management performance between the pre- and post-test ($F_{1,39}=0.141$; $p=0.710$). There was, however, an effect approaching statistical significance with regard to pilots' performance between the trained and untrained group ($F_{1,39}=2.900$; $p=0.097$). The group that had received ADM training tended to perform better than the group that had not received training. There was also an interaction term verging on significance ($F_{1,39}=3.266$; $p=0.078$). The group that received ADM training showed greater gains in performance in the second trial compared to the untrained group.

Table 2
Means and Standard Deviations in performance scores in the Recognition-primed decisions scenario, broken down by both main effects (pre-test/post-test: trained/untrained) on the measures of situation awareness and risk management

Recognition-primed decisions		Group	N	Mean	Standard deviation
Situation assessment	Pre-test	Trained	21	5.43	1.12
		Untrained	20	5.55	1.23
	Post-test	Trained	21	6.10	0.94
		Untrained	20	5.20	1.44
Risk management	Pre-test	Trained	21	5.29	1.19
		Untrained	20	5.30	1.13
	Post-test	Trained	21	5.86	0.73
		Untrained	20	4.95	1.19

Response Selection Decisions. There was an effect approaching statistical significance with regard to pilot performance between the pre- and post-test on the dimension of situation assessment ($F_{1,39}=3.520$; $p=0.068$). This suggested that pilots' situation assessment was rated as having improved on the second trial regardless of whether they received training or not (see table 3). There was also an effect verging on statistical significance between the trained and untrained group ($F_{1,39}=3.277$; $p=0.078$). The group that had received ADM training tended to outperform the group that had not received training. There was no significant interaction effect ($F_{1,39}=1.461$; $p=0.234$). There was no significant difference on risk management performance between the pre- and post-test ($F_{1,39}=2.0641$; $p=0.112$). There was a result approaching statistical significance on risk management performance between the trained and untrained group ($F_{1,39}=4.022$; $p=0.052$). The group that had received ADM training tended to exhibit better performance than the group that had not received training. There was also a significant interaction term between the trained/untrained group and pre-test post-test

trial ($F_{1,39}=5.591$; $p=0.023$). The group that had received ADM training showed greater gains in risk management performance in the second trial compared to the untrained group.

Table 3
Means and Standard Deviations in performance scores in the response selection decisions scenario, broken down by both main effects (pre-test/post-test: trained/untrained) on the measures of situation awareness and risk management.

Response selection decisions		Group	N	Mean	Standard deviation
Situation assessment	Pre-test	Trained	21	5.14	1.46
		Untrained	20	4.75	1.55
	Post-test	Trained	21	5.90	0.99
		Untrained	20	4.90	1.78
Risk management	Pre-test	Trained	21	4.86	1.01
		Untrained	20	4.85	0.99
	Post-test	Trained	21	5.67	0.86
		Untrained	20	4.70	1.17

Resource Management Decisions. There was a significant difference in pilots' situation assessment performance between the pre- and post-test ($F_{1,39}=4.914$; $p=0.033$). Pilots' performance was superior on the second trial (table 4). There was, however, no significance between the trained and untrained group ($F_{1,39}=1.767$; $p=0.191$) and there was also no significant interaction ($F_{1,39}=1.238$; $p=0.273$). Overall, there was an effect verging on significance in risk management performance between the pre- and post-test measures ($F_{1,39}=3.035$; $p=0.089$). Pilots' risk management performance was superior on the second trial. There was no significant difference between the trained and untrained group ($F_{1,39}=0.052$; $p=0.820$) and there was no significant interaction term ($F_{1,39}=2.247$; $p=0.142$).

Table 4
Means and Standard Deviations in performance scores in the resource management decision scenario, broken down by both main effects (pre-test/post-test: trained/untrained) on the measures of situation awareness and risk management.

Resource management decisions		Group	N	Mean	Standard deviation
Situation assessment	Pre-test	Trained	21	4.95	1.56
		Untrained	20	4.80	1.32
	Post-test	Trained	21	5.86	1.15
		Untrained	20	5.10	1.51

Risk management	Pre-test	Trained	21	4.71	1.19
		Untrained	20	4.95	1.00
	Post-test	Trained	21	5.38	1.07
		Untrained	20	5.00	1.52

Non-diagnostic Procedural Decisions. Overall, there was no difference in situation assessment performance between the pre- and post-test ($F_{1,39}=1.007$; $p=0.322$). There was an effect verging on significance in performance between the trained and untrained group ($F_{1,39}=3.593$; $p=0.065$). The group that had received ADM training tended to outperform the group that had not received training (table 5). There was also a significant interaction term between the trained/untrained group and pre-test/post-test trial ($F_{1,39}=19.540$; $p=0.000$). The group that had received ADM training showed significantly greater gains in situation assessment performance in the second trial. There was no significant difference in risk management performance between the pre- and post-test ($F_{1,39}=0.067$; $p=0.797$). There was also no significant difference between the trained and untrained group ($F_{1,39}=1.887$; $p=0.177$). There was a result verging on significance in the interaction term between the trained/untrained group and pre-test/post-test trial ($F_{1,39}=3.266$; $p=0.078$). The group that had received ADM training showed greater gains in performance in the second trial compared to the untrained group.

Table 5
Means and Standard Deviations in performance scores in the non-diagnostic procedural decision-making scenario, broken down by both main effects (pre-test/post-test: trained/untrained) on the measures of situation awareness and risk management.

Non-diagnostic procedural decisions		Group	N	Mean	Standard deviation
Situation assessment	Pre-test	Trained	21	5.00	1.30
		Untrained	20	5.30	1.22
	Post-test	Trained	21	6.19	1.12
		Untrained	20	4.55	1.64
Risk management	Pre-test	Trained	21	4.95	1.16
		Untrained	20	5.25	1.07
	Post-test	Trained	21	5.71	0.96
		Untrained	20	4.60	1.47

Creative problem-solving. There was a significant difference in situation assessment performance between the pre- and post-test measures ($F_{1,39}=10.320$; $p=0.003$). It showed that pilots' performance was better on the second trial than the first trial (table 6). There was no significance between the trained and untrained group ($F_{1,39}=0.187$; $p=0.668$) and there was also no significant interaction term ($F_{1,39}=2.393$; $p=0.130$). There was a significant difference on the dimension of risk management ($F_{1,39}=5.885$; $p=0.020$). It indicated the pilots' performance on risk

management was superior on the second trial. There was no significant difference between the trained and untrained group ($F_{1,39}=0.162$; $p=0.690$). There was also no significant interaction term between the trained/untrained group and trial ($F_{1,39}=2.509$; $p=0.121$).

Table 6

Means and Standard Deviations in performance scores in the Creative problem-solving scenario, broken down by both main effects (pre-test/post-test: trained/untrained) on the measures of situation awareness and risk management

Creative problem-solving		Group	N	Mean	Standard deviation
Situation assessment	Pre-test	Trained	21	4.71	1.35
		Untrained	20	4.90	1.48
	Post-test	Trained	21	5.71	1.01
		Untrained	20	5.25	1.02
Risk management	Pre-test	Trained	21	4.71	1.35
		Untrained	20	4.95	1.76
	Post-test	Trained	21	5.67	0.97
		Untrained	20	5.15	1.23

Discussion

Overall, the results show gains being made in terms of both the participants' situation assessment and risk management skills that are attributable to the short decision making training course. Perhaps the most direct indication of the efficacy of the ADM training course lies in the significant interaction effects obtained. These interaction terms indicate disproportionate gains in performance on the second trials (post ADM training) in the participant group that received ADM instruction. To summarize, significant results (or results approaching significance) were obtained showing improvements in participant's performance in the scenarios concerned with go/no go decisions, recognition-primed decisions, and non-diagnostic procedural decisions. With regard to risk management, significant results (or results verging on significance) were observed in the go/no go decision making scenario, recognition-primed decision making scenario, response selection, and non-diagnostic procedural decision making scenario. These results are summarized in table 7.

Even though every effort was made to ensure that the pre- and post-training decision making scenarios were of equivalent difficulty, inspection of the results from the untrained group would suggest that in several cases the post-test scenarios were actually slightly more difficult (see tables 1, 2 and 5). Nevertheless, in spite of this evidence that would suggest that these post- training scenarios were more difficult, the trained group still generally showed improvements in situ-

ation assessment and risk management performance (see the associated interaction terms). In all cases, the performance of the group that received the ADM training course improved.

Table 7

The summary of main effects and interaction effects of paper-pencil trials on both dimensions of situation assessment and risk management across six basic types of decision-making scenarios

Six basic types of decision-making	Dimensions of evaluation	Main effect of before/after training	Main effect of trained/untrained	Interaction effects
Go/no go decisions	SA		👍	✓
	RM			👍
Recognition-primed decisions	SA			✓
	RM		👍	👍
Response selection decisions	SA	👍	👍	
	RM		👍	✓
Resource management decisions	SA	✓		
	RM	👍		
Non-diagnostic procedural decisions	SA		👍	✓
	RM			👍
Creative problem-solving	SA	✓		
	RM	✓		

Note: 👍 indicates a result approaching significance ($p < 0.10$); ✓ Indicates a significant result ($p < 0.05$); SA = Situation Assessment; RM= Risk Management.

For the evaluation of both Situation Assessment and Risk Management performance in the narrative answers produced, a. The steps that should be undertaken and sources of information that should be interrogated in each circumstance were listed, these being factors underlying Situation Assessment performance in particular. Emphasis on the risk management dimension was placed upon the generation and analysis of options and the quality of reasoning underlying the pilot's final decision based specifically on the control of risk.

The results obtained add support to the findings of earlier research (e.g. Buch and Diehl, 1984; Connolly, Blackwell & Lester, 1989; Endsley, 1993; Klein, 1993 & 1997; Orasanu, 1993; Prince & Salas, 1997) that suggested that ADM was trainable. Orasanu (1993) advocated there was no evidence that generic training techniques to improve decision making skills would be effective as different component skills were involved when making different basic types of decisions. As a result of this Li & Harris (2005) elicited the opinions of a large sample of instructor

pilots concerning the best ADM mnemonic-based methods for use in a variety of different types of flight situations. SHOR (Wohl, 1981) was identified as potentially the best ADM mnemonic in a time-limited situation; DESIDE (Murray, 1997) was rated as being superior for more complex, knowledge-based decisions where more time was available. The results obtained in this study support the conclusions of the earlier opinion survey. These decision making mnemonic-based methods promote better ADM. There is now empirical evidence demonstrating that pilots trained in the use of these techniques actually produce superior performance on two of the essential components underlying ADM for at least some varieties of decision making problems.

The data in the narrative reports produced by the participants in each decision making scenario suggested that the majority of pilots who had received ADM training applied the most appropriate ADM mnemonic method for a given situation. The SHOR mnemonic tended to be applied in the go/no-go decision making scenario, recognition-primed decision-making scenario and in the response selection decision-making situation. DESIDE was most commonly used in the remaining scenarios (resource management decisions, non-diagnostic procedural decisions and creative problem-solving).

Conclusions

This research investigated the efficacy of a short ADM training course using two mnemonic-based methods (SHOR and DESIDE) to improve ROC Air Force pilot decision-making in six different basic types of decision-making scenarios. The results from simple paper-and-pencil based evaluations assessing the knowledge acquired show that such a short training course is generally effective in improving pilots' situation assessment and risk management skill (two underpinning requirements for effective decision-making) in a range of decision-making situations. Complementary research undertaken in a flight simulator has also shown behavioral gains in decision making by those who underwent the training course (Li & Harris, 2006). These complimentary behavioral gains further establish the validity of the use of pencil and paper based tests to evaluate the ADM training course. They provide convergent evidence to support the efficacy of the decision making training program. However, the longer-term effectiveness of such courses needs evaluation to see if it translates into improved decision-making behavior during day-to-day operations, which ultimately results in a reduction in the accident rate attributable to poor decision-making. By necessity, the initial evaluations of the training program focused upon 'problem' situations where pilots were required to make a satisfactory decision to avoid a potential accident. Further research is required to establish if the ADM principles conveyed in the training course are equally as successful in lower workload, less pressured decision making situations. Nevertheless, this simple, short, cost-effective training program in the appropriate use of ADM mnemonic methods can potentially produce significant gains in flight safety. Such a course may easily be integrated into the existing CRM and/or simulator-based training programs currently undertaken by cadet pilots in the ROC Air Force. Furthermore, there is no reason why a

modified version of the ADM training course devised should not be equally as successful in a civil aviation training organization.

References

- Baron, J. and Hershey, J.C. (1988). Outcome bias in decision evaluation, *Journal of Personality and Society Psychology*, 54, 569-579.
- Brown, R.V., Kahr, A.S., & Peterson, C (1974). *Decision analysis for the manager*. New York: Holt, Rinehart & Winston.
- Buch, G. and Diehl, A. (1984). An investigation of the effectiveness of pilot judgment training, *Human Factors*, 26(5), 557-564.
- Connolly, T.J., Blackwell, B.B. and Lester, L.F. (1989), A Simulator-based approach to training in aeronautical decision making, *Aviation, Space, and Environmental Medicine*, 60(1), 50-52.
- David, G. (1997), Decision making training for aircrew. In R. Flin, E. Salas, M. Strub, & L. Martin, (Eds.), *Decision Making Under Stress: Emerging Themes and Applications* (pp. 243-251). Aldershot, England: Ashgate.
- Endsley, M.R. (1997). The role of situation awareness in naturalistic decision making. In C.E. Zsombok & G. Klein, (Eds.), *Naturalistic Decision Making* (pp. 269-284). Lawrence Erlbaum, Mahwah.
- Endsley, M.R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology* 3(2). 157-168.
- Endsley, M.R. and Bolstad, C.A. (1994). Individual differences in pilot situation awareness. *International Journal of Aviation Psychology*, 4(3). 241-264.
- Federal Aviation Administration (1991). *Aeronautical decision-making*, Report no. Advisory Circular 60-22, US Department of Transport, Washington.
- Fischer, U., Orasanu, J., and Wich, M. (1995). Expert pilots' perceptions of problem situations. In R.S. Jensen, & L.A. Rakovan (Eds.), *Eighth International Symposium on Aviation Psychology*, Vol. 2. (pp. 777-782). Columbus, OH, U.S.A.: The Ohio State University.
- Hormann, H.J. (1995). FOR-DEC: a perspective model for aeronautical decision making. In R. Fuller, R. Johnston, & N. McDonald (Eds.), *Human Factors in Aviation Operations* (pp. 17-23). Aldershot, England: Ashgate Publishing.
- Hunter, D.R. (2003). Measuring general aviation pilot judgment using a situational judgment technique, *International Journal of Aviation Psychology*, 13(4), 373-386.
- Janis, I.L. and Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.
- Jensen, R.S. (1997). The boundaries of aviation psychology, human factors, aeronautical decision making, situation awareness, and crew resource management. *International Journal of Aviation Psychology*, 7(4), 259-268.
- Jensen, R.S. (1995). *Pilot judgment and crew resource management*, Aldershot, England: Ashgate.
- Jensen, R.S., Guilke, J. and Tigner, R. (1997). Understanding expert aviator judgment. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision Making Under Stress: Emerging Themes and Applications* (pp. 233-242). Aldershot, England: Ashgate.
- Kaempf, G.L. & Orasanu, J. (1997). Current and future applications of naturalistic decision making. In C.E. Zsombok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 81-90). Mahwah: Lawrence Erlbaum.
- Kirkpatrick, D.L. (1998). *Evaluating training programs*. San Francisco, CA: Berrett-Koehler.

- Kirkpatrick, D.L. (1976). Evaluation of training. In R.L. Craig & L.R. Bittel (Eds.), *Training and Development Handbook* (pp. 18.1-18.27). McGraw Hill, New York.
- Klein, G. (1997). The current status of the naturalistic decision making framework. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision Making Under Stress: Emerging Themes and Applications* (pp. 11-28). Aldershot, England: Ashgate.
- Klein, G.A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsombok (Eds.), *Decision Making in Action: Models and Methods* (pp. 138-147). Norwood, New Jersey: Ablex.
- Klein, G.A. (1993). Twenty questions: Suggestions for research in naturalistic decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsombok, (Eds.), *Decision Making in Action: Models and Methods* (pp. 389-403). Norwood, New Jersey: Ablex.
- Klein, G.A. & Woods, D.D. (1993). Conclusions: Decision making in action. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsombok (Eds.), *Decision Making in Action: Models and Methods* (pp. 404-411). Norwood, New Jersey: Ablex.
- Li, W.C. & Harris, D. (2005). Aeronautical decision-making: Instructor-pilot evaluation of five mnemonic methods. *Aviation Space Environmental Medicine*, 76, 1156-1161.
- Li, W.C., Harris, D., and Yu, C.S. (2005a). The identification of training needs for developing aeronautical decision-making training programs for military pilots. In R.S. Jensen (Ed.), *Thirteenth International Symposium on Aviation Psychology*, (pp. 351-356). Oklahoma City, OK: The Ohio State University.
- Li, W.C., Harris, D., and Yu, C.S. (2005b). The investigation of suitability of aeronautical decision-making mnemonics in tactical environments. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 2187-2191). Orlando, FL.: Human Factors and Ergonomics Society.
- Li, W.C., and Harris, D. (2006, October). SHOR and DESIDE: Evaluating the effectiveness of ADM training using a flight simulator. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, CA.
- Maher, J. (1989). Beyond CRM to decisional heuristics: An airline generated model to examine accidents and incidents caused by crew errors in deciding. In R.S. Jensen (Ed.), *Fifth International Symposium on Aviation Psychology*, (Vol. 2, pp. 439-444). Columbus, Ohio: Ohio State University.
- Murray, S.R. (1997). Deliberate decision making by aircraft pilots: A simple reminder to avoid decision making under panic. *International Journal of Aviation Psychology*, 7(1), 83-100.
- Noble, D. (1993). A model to support development of situation assessment aids. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsombok (Eds.), *Decision Making in Action: Models and Methods* (pp. 287-305). Norwood, New Jersey: Ablex.
- O'Connor, P., Flin, R., Fletcher, G. & Hemsley, P. (2002). Methods used to evaluate the effectiveness of flightcrew CRM training in the UK aviation industry, *Human Factors and Aerospace Safety*, 2(3), 235-255.
- O'Hare, D. (2003). Aeronautical decision making: Metaphors, models, and methods. In P.S. Tsang, & M.A. Vidulich (Eds.), *Principles and Practice of Aviation Psychology: Human Factors in Transportation*, (pp. 201-237). New Jersey: Lawrence Erlbaum.
- O'Hare, D. (1992). The artful decision maker: A framework model for aeronautical decision making, *The International Journal of Aviation Psychology*, 2(3), 175-192.
- Oldaker, I. (1996). Pilot decision making - An alternative to judgement training, *Technical Soaring*, 20(2), 36-41.
- Orasanu, J. (1993). Decision making in the cockpit. In E.L. Wiener, B.G. Kanki, & R.L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 137-172). San Diego, CA: Academic Press.

- Orasanu, J. (1997). Stress and naturalistic decision making: Strengthening the weak links. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.). *Decision Making Under Stress: Emerging Themes and Applications* (pp. 43-66). Aldershot, England: Ashgate.
- Prince, C. and Salas, E. (1997). Situation assessment for routine flight and decision making, *International Journal of Cognitive Ergonomics*, 1(4), 315-324.
- Reason, J. (1990). *Human error*. New York, NY: Cambridge University Press.
- Wohl, J.G. (1981), Force management decision requirements for air force tactical command and control, *IEEE Transactions on Systems, Man and Cybernetics*, SMC-11(9), 618-639.

Book Reviews

Developing Strategies for the Modern International Airport East Asia and Beyond

by Allan Williams

Reviewed by
John C. Di Renzo Jr.
*Oklahoma State University
Stillwater, OK*

This book is based on the premise that the aviation industry has a decisive role in the economic and social development of the modern international economy. The author, Allan Williams, achieves his stated objective by exploring ideas for further development and research. The book is both thorough and instructional in nature and is certainly written for the practitioner, but it is also of value to aviation professionals interested in strategic planning and development within the aviation industry. It is an excellent review of past and current aviation industry issues. The first several chapters primarily discuss issues in the aviation industry and clearly articulate a number of divergent views on concepts such as "Globalization" and its impact on strategic decision making in the aviation industry. After an exhaustive discussion of economic, geopolitical, regulatory, historical aviation and business issues, the author makes a more concerted effort to discuss the issues in the context of aviation development in the East Asia market. East Asia, not only includes East Asia, but also Northeast Asia and China. The book specifically addresses strategic concerns in places such as China, Japan, Hong Kong, South Korea, Malaysia, Thailand, and Singapore.

The primary focus of this book is to identify and analyze the role of the international airport as an increasingly multifunctional agency and multi-modal mass transit system. That is, as an agency providing a number of services in addition to air transportation and as an agency offering multiple modes of transportation as urban sprawl makes new demands especially in rapidly expanding East Asian economies. The author thoroughly develops complex issues and does a good job of fairly discussing divergent views; however, he makes no attempt to simplify this complexity for the reader. He presents the issues with all their rich complexity and often leaves conclusions to the reader. His discussion of "Globalization" in Chapter 2 illustrates this point.

Key aviation related topics developed in the first three chapters include deregulation, privatization, open skies agreements, liberalization, Low-Cost Carrier (LCC) competition, hub-and-spoke development, route competition, predatory pricing, and other issues. At the same time, however, Williams discusses business, economic and political issues not directly affiliated with aviation, but nonetheless, having a profound impact on aviation development and future strategy. These topics include, but are not limited to the creation of free trade zones, supply chain management, the emergence of the ubiquitous international enterprise, foreign direct investment, competitive advantage and other modern business and economic theories and concepts. An example is his reference to Michael Porter's five competitive forces model as applied to the aviation industry. He is obviously aware that all readers may not understand these multidisciplinary references and, in a note, directs the reader to other sources for explanation. Consequently, the reader should have a great breadth of knowledge regarding not only aviation industry topics, but also business, economic, political, and even information technology subject matter.

In the beginning of the text and throughout the book Williams questions whether deregulation of the airlines and liberalism actually allowed competition and thereby, diminished the role of the state. He suggests neither the supporters nor the critics of deregulation got it right. Aviation grew after deregulation largely as a function of internationalization of trade and commerce, not necessarily because of market liberalization within the airline industry. His key point, which is reinforced throughout the text, is that one cannot understand future aviation development without first understanding key business and economic drivers in the emerging global economy.

The remainder of the book focuses more directly upon East Asia aviation market development in the context of the concepts developed in previous chapters. Williams reminds the reader that the prevailing expert opinion is that the Asian-Pacific rim economies will emerge as market leaders in the aviation industry. This is for a number of geopolitical and economic reasons and is so despite the downturn of the East Asian economies (i.e. the "Little Tigers") in 1997. At the core of these developments are China and the states that comprise the Association of Southeast Asia Nations (ASEAN). Williams compares and contrast development of East Asian economies with those of the western countries. He duly notes that while several states such as China, Japan, Singapore, Hong Kong, Thailand, and Malaysia are already global competitors in the aviation industry, there is nonetheless great disparity among other less economically developed states.

In a compelling discussion, Williams reviews the influence of rapid urban growth in East Asian cities and discusses its impact on airport development. Key factors are the demand for aviation services and the ever-increasing importance of logistics and supply chain management development. Other issues include peripheral development at the edge of metropolitan centers and the quality of Foreign Direct Investment (FDI) by multi-national companies. Clearly, the less developed economies of the region are affected by this investment. He refers to this phenomenon as an "aerotropolis" characterized by multi business dynamics in rapidly growing cities such as Beijing and Shanghai. Additionally, he addresses the role of major hub airports as crucial to the development of multi-modal transportation networks that support urbanization in such cities.

Other issues related to economic growth in East Asia include the strategic importance of airfreight, LCCs in East Asia, and regulatory problems limiting FDI in East Asia. Most notable is his discussion of a “speculative” view of emerging problems in East Asia. These themes include a number of significant technological, managerial, and operational problem areas, which are matters of great controversy within the industry. Examples include the prospects for near-term global economic growth, the price of oil, China’s undervalued currency, the longevity of the legacy carriers, and the development of ultra long-haul aircraft providing non-stop services for business and leisure travelers.

The flow of ideas in each chapter transitions well from one topic to the next as Williams develops and refines varied themes. However, the discussion occasionally seems somewhat redundant because previous topics are developed in a slightly different context throughout the book. The book is relatively short and easy to read for those fairly well-versed with issues in business and the airline industry.

The author completed exhaustive research as is evidenced by numerous bibliographic entries. All bibliographic references appear to be relevant and most are from reasonably recent publications.

The book is most appropriate for aviation professionals in government and industry, aviation regulatory agencies, airport management personnel, and faculty and students seeking degrees in aviation and/or business related disciplines. The author’s multidisciplinary approach assumes a more sophisticated reader who is well-versed in a number of subject areas. What is most compelling, however, are the many unanswered questions posed by Williams. This book would certainly form a basis for a more detailed examination of the many inter-related and interdisciplinary variables discussed and, therefore, would be of great value to investigators in varied fields of research.

What We Should Know About Human Error: A Review of Ten Questions about Human Error

by Sidney Dekker

Reviewed by
Todd P. Hubbard
*Oklahoma State University
Stillwater, OK*

Sidney Dekker has been busy identifying and describing human error for the better part of a decade. However, *Ten Questions about Human Error* is perhaps one of his crowning achievements to date. What makes this read different from some of his other books and research articles is that the pages of this 10-chapter book are packed with philosophical and psychological subplots. After reading the Preface, my first impression of the narrative was that the word choice and intellectual engagement made the experience delicious; which I agree appears to be an odd first impression. However, when one's intellectual appetite is stimulated by the seasoning of psychosocial and philosophical descriptions of human performance as they apply to the determination of when and how errors are made, one is compelled to turn off the cell phone and retreat into a quiet space to feast on this extraordinary book.

When psychologists were invited by NASA and the federal government (the Federal Aviation Administration and the National Transportation and Safety Board in particular) to examine the world of air carrier operations in the 1970s, they brought with them the theories and conceptual thinking relevant to each area of interest within the overarching domain of psychology. To psychologists, the well-ordered work environment of the airline pilot provided a perfect experimental workspace where behavior could be observed and cataloged. To pilots and others within aviation, what the psychologists said was happening did not always ring true with those with flight experience. Dekker tells us why psychologists perceived flight operations as they did. He also tells us that psychological viewpoints are not shared across the discipline and that perceptions owned by psychologists continue to evolve.

Dekker does a masterful job describing the differences between the branches of psychology that are the most prevalent in aviation research. He exposes the weaknesses of experimental methods used by behaviorists, tracing these weak-

nesses back to the beginning of experimental psychology, mostly centered on the Würzburger School.

Perhaps the most important achievement in this text is the clear and determined revision of traditional definitions or descriptions of human error. Dekker starts by examining the Newtonian-Cartesian preoccupation with philosophical dualism, which suggests that all phenomena are perceived to be of two categories: the mind (*res cogitans*) or the material world (*res extensa*). Human error, according to the dualistic view, can be described as one part human and one part material failure. If one can eliminate the material failure, then the only other likely suspect is the human in the loop. However, as Dekker so craftily presents, not all judgments of error-making can be so easily made, when one examines the work environment, the persons operating within that environment, and the organizational influence that applies pressure to the human participant within the work environment.

In an attempt to uncover the roots of ill-conceived experimental methods that would lead to errors in observation of participants in research studies, Dekker takes us back to what he claims is the point where psychologists took the wrong direction in their study of human behavior. Instead of watching human participants complete tasks in their normal environment and at the normal pace, and without interference by observers, experimental psychologists following the Würzburger School paradigm subdivided the working environment into isolated sections, allowing researchers to examine smaller portions of behavior. This led psychologists to draw conclusions about human behavior and on participant sense-making that were not tied to the dynamic environment—the sociotechnical—but a pseudo-environment, where natural occurrences were blanked out.

Dekker believes that the traditional approach to assessing human error is taking us all in the wrong direction and that this traditional approach will actually have a detrimental affect on system safety. He advocates examining pilots and others in their natural environment, while they are performing tasks naturally, and are behaving in ways that incorporate the affects of organizational influence, crew influence, and operational environmental influence. He has created a compelling argument for resisting the blame game and instead creating an environment where finding system errors is more important than victimizing or criminalizing humans that make errors. Error-making is what comes natural to humankind. Why should this behavior be treated as an oddity, an abnormal event?

I use this text in my Human Factors course at Oklahoma State University, because it presents the alternative argument for human error, which is so badly needed in 21st century classrooms. Students think more critically if they are presented with alternative views, not when they are presented only one view. I use it as a companion text, along with more traditional treatments of human factors in aviation.

Collegiate aviation undergraduate readers might find some of the terms and concepts to be unfamiliar, especially if student does not have any background in psychology or philosophy. Much of the literature that collegiate aviation students read is highly technical, pragmatic, and structured, perhaps over excising the left

hemisphere while denying the more imaginative right hemisphere to engage. Despite the vocabulary, I would still recommend the text for undergraduate courses in human factors, because Dekker spends a great deal of time describing these difficult terms and concepts in each of his chapters. In fact, this is one of the charms of the book. The author introduces a concept and then incrementally explains the many facets of the concept, from several perspectives—not all of which are necessarily in agreement with those of the author. In this regard, the text is an equal opportunity treatment of psychological issues from across decades of inquiry.

Beyond the undergraduate reader, I would also recommend the text for graduate students, human factors educators and scientists, aviation psychology educators and scientists, air carrier operations managers, collegiate flight program managers and certified flight instructors, safety officers and managers, Federal Aviation Administration Aviation Safety Inspectors (pilot and maintenance), and accident investigators with the National Transportation and Safety Board. I believe that Dekker's message should be part of every aviation-training program.

Lastly, I wish to draw attention to the intellectual value of this text. Scholars and scientists might not give credit to airplane pilots and maintainers for having any measure of intellect or scholarly sophistication. These people might also consider any scholarly dialogue among pilots to be of lesser value, based on this bias. Sidney Dekker destroys this bias, not only because he is a pilot, a scholar, an intellectual, and a brilliant cognitive engineer, but also because he can command respect across so many disciplines, without pretense. If you want to stretch your mind and examine your thoughts about human error and system safety, you will want to read this book.