2006 International Journal of Applied Aviation Studies Volume 6, Number 1 A publication of the FAA Academy

International Journal of Applied Aviation Studies

A Publication of the FAA Academy Oklahoma City, Oklahoma

Volume 6, Number 1, 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-530
PO Box 25082
Oklahoma City, OK 73125

International Journal of Applied Aviation Studies
Volume 6, Number 1
ISSN Number: 1546-3214

1st Printing June 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.

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
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
David Conway	Southeastern Oklahoma State University, Durant, OK
Janet 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
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
Rebecca Lutte	University of Nebraska at Omaha, Omaha, NE
Beth Lyall	Research Integrations, Inc., Tempe, AZ

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
John C. Morey	Dynamics Research Corporation, Andover, MA
Kathleen Mosier	San Francisco State University, San Francisco, CA
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
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
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 III	Rotary-Wing Aviation Research Unit, 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
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.

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

EDITOR'S NOTES

Quality in airline safety must be a goal. Our lead article by Stolzer, Wu, and Halford focuses on the airline industry's need for an efficient quality approach to improving safety. The authors introduce a technique in which current Flight Operations Quality Assurance (FOQA) methodology can be enhanced with the sophisticated quality and statistical concepts found in Six Sigma. The authors provide an overview to both FOQA and Six Sigma and provide an interesting hypothetical exemplar case study using Six Sigma methodology on a FOQA problem.

In our second article, Bass presents a comparison of four quantitative methods of measuring human judgment performance in a simulated horizontal air traffic conflict prediction task. Four quantitative methods for modeling human judgment performance in unsure environments (signal detection theory, fuzzy signal detection theory, lens modeling, and skill scores) were compared. In addition to individual differences, gender, noise level, and session (experience) effects and their two-way interactions were investigated.

The Landry and Jacko article describes a method for leveraging an existing design methodology – procedure context – to provide guidance on the content of displays for procedure following. This paper outlines a method of identifying information which, when displayed dynamically, benefits pilots executing an instrument approach procedure. The article provides a guide for those wishing to design support for procedure following.

Does the everyday experience of flying or flight instruction together with recent flight experience and flight review requirements, eliminate the need for ongoing study or rehearsal of aeronautical knowledge? Casner, Heraldez, and Jones article focuses on the retention of aeronautical knowledge. Four experiments were conducted. The results support some hypotheses but also further demonstrate that there are no simple-to-measure determinants of what aeronautical knowledge will be remembered and what will be forgotten.

Kristovics, Mitchell, Vermeulen, Wilson, and Martinussen's article presents the findings from the four-factor Aviation Gender Attitude Questionnaire (AGAQ). The questionnaire gathered data from 2009 pilots from the USA, South Africa, Australia, and Norway. The authors analysed the data for possible gender bias among pilots. The results of the four-factor AGAQ are discussed along with the benefits, and the limitations, of the measure.

This article by Casner is a follow-up to the author's 2005 article published in the International Journal of Applied Aviation Studies 5(1). This article demonstrates how pilots can overcome the loss of navigational awareness when using the GPS and moving map display. In this follow-up study, eight pilots used GPS and moving map displays to navigate between the same circuit of checkpoints while performing one additional task. The performance of this additional task provided a significant advantage in alleviating the "out-of-the-loop" phenomenon associated with using GPS and moving map displays.

Can the introduction of Technically Advanced Aircraft (TAA) in ab-initio training be accomplished successfully? Dahlström, Dekker, and Nählinder interviewed three flight instructors responsible for the introduction of TAAs in ab-initio training at a flight school and administered questionnaires to instructors and to ab-initio students. The results are presented and concerns discussed. Further study is recommended

Dillman and Lee's purpose is not to determine the effectiveness or validity of the situational judgment test (SJT), but illustrate the utilization of the situational judgment test within a collegiate flight-training program. The authors propose that the more flight students are introduced to the process of aeronautical decision-making, the stronger their understanding of the limitations of their abilities and devices, such as the SJT enhance their decision-making processes.

McDermott's article offers the results of a study of pilot instrument proficiency and how computer-based flight simulation can improve pilot performance. Although practicing instrument approach procedures in simulation is not a requirement for recent flight experience under Federal Aviation Regulation 61.57, the author's study proves that practice via computer-based flight simulation is a valuable resource that pilots should consider when maintaining, and improving, their instrument flight skills.

Wetmore and Lu investigate the relationship between aeronautical decision-making (ADM) and crew resource management (CRM) skills in fatal general aviation accidents and the effects of hazardous attitudes. Hazardous attitudes are shown to have a devastating effect on evaluating risk, making decisions, and utilizing all available resources, three of the most important CRM skills. The authors recommend that all pilots, at every level of pilot certification, should receive better ADM and CRM training to reduce the effects of hazardous attitudes.

Hamilton focuses on the effective behavior that is conducive to team building and solidarity in the air carrier cockpit. A multidimensional scaling methodology was applied to determine a classification of constructive leader behaviors and their relationship to the exercise of power.

KC

International Journal of Applied Aviation Studies

Volume 6, Number 1 2006

Formal Papers

Book Review

International Journal of Applied Aviation Studies, Volume 6, Number 1

Formal Papers

Six Sigma Applied to Flight Operations Quality Assurance: An Exemplar Case Study

Alan J. Stolzer
Saint Louis University
3450 Lindell Blvd.
St. Louis, Missouri 63103
(314) 977-8203
stolzeraj@slu.edu

Han Wu
SAGEM Avionics, Inc.
2701 Forum Drive
Grand Prairie, TX 75052-7099
(972) 314-3600
han.wu@sagemavionics.com

Carl Halford
Saint Louis University
3450 Lindell Blvd.
St. Louis, Missouri 63103
(314) 977-8203
carl.halford@gmail.com

Abstract

Due to the requirement to maintain and improve the safety record of commercial air transportation in the United States (U.S.) despite increasing traffic, several proactive safety programs have been introduced in recent years. Among these proactive safety programs is a form of Flight Data Monitoring (FDM) known in the U.S. as Flight Operational Quality Assurance (FOQA). FOQA is a program utilizing quantifiable, objective data collected from the air carrier aircraft's data recording system. The data is then analyzed to identify trends and other indicators of potential safety problems. With few exceptions, FOQA data analysis has been rudimentary, often limited to relatively simple statistical methods. The purpose of this study was to introduce a method in which current FOQA methodology can be enhanced with the more sophisticated quality and statistical concepts found in Six

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.

Sigma – a structured, data-driven approach built upon to eliminating defects through the reduction of variation in processes. A general introduction to both FOQA and Six Sigma is provided, along with a hypothetical exemplar case study using Six Sigma methodology on a FOQA problem, i.e., tail strikes during takeoff.

The U.S. air transportation system is considered one of the safest forms of transportation in the world (NASA, 2004). Airline safety departments have developed and implemented numerous proactive safety initiatives over the past several years such as the Advanced Qualification Program (AQP), Flight Data Monitoring (FDM), Aviation Safety Action Program (ASAP), Internal Evaluation Program (IEP), and the Voluntary Disclosure Reporting Program (VDRP), with the primary intent to improve safety. However, additional gains may be possible by implementing a widely utilized and highly regarded quality program known as Six Sigma. This research provides an overview of one of the most significant proactive safety, airline-oriented flight data monitoring programs - Flight Operations Quality Assurance (FOQA), and Six Sigma. With that background established, an exemplar case study of the application of Six Sigma principles to a FOQA problem is then presented.

Airline Safety

Throughout most of the aviation industry's history, the primary method of research concerning the mitigation of risk has been reactive, that being post-event analyses of incidents and accidents. Many significant advances in safety have resulted from this methodology: decreases in serious wake turbulence encounters due to greater in-trail spacing, improved cargo compartment smoke detection systems, transponder-based intruder conflict alerting systems, improved windshear detection systems at airports, to name but a few. The list of advances is long indeed, and proves the worth of rigorous post-accident investigation (NTSB, 2004). However, by the 1990s, investigators and regulators alike were coming to the realization that there was a limit to the effectiveness of post-hoc fixes to safety problems, and that limit was based upon relatively simple math.

Figure 1 depicts the accident rates per 100,000 flight hours for U.S. scheduled air carriers operating under 14 CFR 121 from 1985 through 2004 (NTSB, 2005). Although the accident rate is somewhat uneven year to year, the linear fit line (indicated at a value of approximately .22 accidents per 100,000 flight hours) suggests that the rate has stabilized despite an increase in the number of flights. Nevertheless, Weener (1990) hypothesized that if the current rate remained the same, a significant rise in the number of hull losses would occur, thus emphasizing the necessity for proactive safety methodologies such as FOQA.

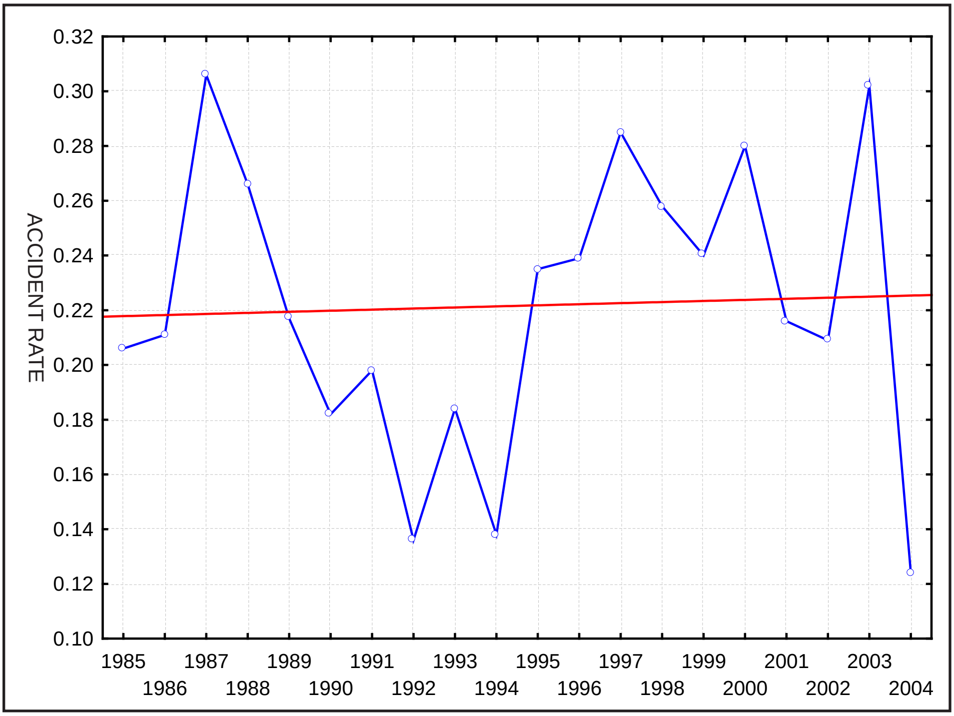


Figure 1. Accident Rates Per 100,000 Flight Hours, 1985 through 2004, for U.S. Air Carriers Operating Under 14 CFR 121, Scheduled Service (Airlines).

Flight Operations Quality Assurance

FOQA, a term coined by the Flight Safety Foundation (FSF) in the early 1990s, is a form of FDM where flight related parameters are collected and analyzed for the purposes of monitoring and improving flight operations with a potential byproduct being the enhancement of flight safety. FOQA methodology has involved:

1. Selecting parameters to monitor and defining events.
2. Capturing, retrieving, and analyzing recorded flight data to determine if the pilot, the aircraft's systems, or the aircraft itself deviated from typical operating norms.
3. Identifying trends or singular anomalies.
4. Taking remedial steps to correct problems.
5. Continuously monitoring the effectiveness of actions taken.

The advantage of data monitoring has been evident due its prevalence in various industries other than aviation. For example, automotive engineers utilize telemetry to monitor multiple aspects of a car's design and performance. Formula One teams feed off telemetry information to determine whether changes made to a car's setup, results in higher performance. Hospitals utilize this technology to monitor patients' health with information logged for the detection of unwanted trends. Complex modern systems such as high-tech manufacturing plants, subway networks, nuclear power plants, and power grids have utilized data monitoring to understand the processes that occur throughout the system.

FOQA is unique among other proactive airline safety initiatives in that it has been the sole utilizer of objective, quantitative data. Depending upon the capabilities of the aircraft involved, FOQA collects parameters from hundreds of sensors located throughout the structure that feed analog and digital input to recording equipment onboard. On a typical Boeing 757 manufactured 15 years ago, for example, 200 to 300 parameters can be recorded and stored each second. Sophisticated airplanes produced today are capable of capturing over 2,000 parameters per second (Phillips, 2002). Pilot control inputs, control surface positions, engine performance parameters, avionics information, and numerous other parameters have been recorded throughout the duration of the flight. FOQA analysts then routinely probe the data to monitor and detect trends in the operation of the aircraft, to determine if exceedances (i.e., an event that exceeds predetermined thresholds) have occurred, and to assess the efficiency of operations. By detecting trends and patterns, it is possible to correct potential problems before they occur.

Using advanced flight data analysis software such as the SAGEM Analysis Ground Station (AGS), FOQA analysts have been able to examine individual flights or aggregated data from numerous flights tracked over time so that statistical trending, through robust reporting and animation modules, can be performed. An aggregate study might examine, for example, the number of unstabilized approaches at a particular airport per month over the last 12 months. This type of analysis provides potentially valuable information, especially in terms of whether the airline's performance is improving, holding constant, or deteriorating. This look at aggregate exceedances over time provides airline managers with a new perspective on potential problems that would not otherwise be apparent. Based on the trend analysis, airline managers can take corrective action to reduce or eliminate detected exceedances by focusing on the root causes and making or recommending changes.

In spite of the availability of both internal and external sources of information coupled with increasingly sophisticated computational technology, many airline managers could gain from additional knowledge and training in the use of quality and statistical tools necessary to reap the maximum advantage from these potent sources of information. In a survey conducted by the GAIN working group, it was revealed that most safety personnel have not received much, if any, formal training directed at the effective use of analytical tools (Global Aviation Information Network, 2001). The report revealed that some sophisticated tools are being used, for example, one airline reported using a tool called Procedural Event Analysis Tool, another reported employing Reason's model and root cause analysis, and several airlines perform flight data analysis and trending using internal databases (Global Aviation Information Network, 2001). What may be most noteworthy regarding the list of tools used is the absence of well-established quality tools and processes such as control charts, Pareto charts, scatter diagrams, cause and effect diagrams, and many other quality management tools (Stolzer & Halford, 2004). FOQA's effectiveness has been determined by the ability of the user to properly determine aspects of a flight operation to be monitored, maximizing the flight data analysis software's potential, formulating analysis results that are meaningful to upper management and other stakeholders (such as pilots), and finally, implementing proper frameworks to remedy and monitor any potentially dangerous trends. The purpose of this work is to assert that FOQA's effectiveness and, thus, airline safety may be enhanced by the application of Six Sigma methods. Six Sigma is a disciplined, data-driven approach to eliminating defects via the reduction of variance. To understand these methods more thoroughly, a rudimentary discussion of distribution, variation, and Six Sigma as a management system is presented.

Six Sigma

In the early and mid-1980s, Motorola engineers developed the concept of Six Sigma – including the standard itself, the methodology and the cultural change associated with it – to provide greater resolution in measuring and decreasing defects. Six Sigma is credited with helping Motorola save billions of dollars by optimizing many processes throughout the company related to manufacturing and other sectors (Motorola, 2004). Inspired by Motorola's success, hundreds of companies around the world have adopted Six Sigma as a way of doing business.

The fundamental objective of the Six Sigma methodology is the implementation of a data-driven strategy that focuses on variation reduction and process improvement through the application of Six Sigma improvement projects. By determining the degree of variation present in an existing process, one has been able to determine its capability by referring to the standard normal distribution, where measures of dispersion can be correlated with probabilities of failure, and parts per million (ppm) defectives.

Distribution

In a standard normal distribution (also known as the "bell curve" or Gaussian distribution), the area under the curve has represented the percentage and thus the probabilities of values contained within and beyond each standard deviation. In fundamental statistics one learns that for a distribution of Mean (μ) = 0 and σ = 1, approximately 68% of values are contained within $\pm 1\sigma$ around the mean, ~96% of the cases within $\pm 2\sigma$ around the mean, and ~99.7% of the cases within $\pm 3\sigma$ around the mean. Therefore, a process capability established at 2σ would result in an acceptable rate of ~96% and a probable "defect" (out of specifications) rate of ~4%; out of every 100 outputs, probability theory states that approximately four will be defective.

Variation

According to Park (2003), the two forms of variation, common cause and special cause, are the primary foes of quality control. Common cause variation is known as naturally occurring variation, where the sources of variation form a stable and repeatable distribution over time. Such a process is 'in control'. Special causes of variation, on the other hand, refers to those instances where an external element causes the overall process distribution to shift erratically causing it to be 'out of control'. For example, if a basketball player with a historical shot percentage of .800 were to attempt 100 in any given day, the conversion rate will naturally and expectedly vary with an ~80% success rate. However, during this process if a special cause is introduced, such as another player attempting to block the shot, this will likely significantly reduce the shooter's ability to convert the free throws. The identification of variation – being able to differentiate between common and special causes – and reduction or elimination of special cause variation are critical elements in ensuring that a process remains standardized and under control.

Prior to the mid-1980s, Motorola was operating at 4σ, but desired a higher standard to account for variations in the process over time (Harry, n.d.). Motorola engineers determined that once a long-term process is no longer centered at the specified target (design specification) due to a variation of ±1.5σ, the rate of defects increases and the capability to produce results within specifications is hampered. This results in a process at 2.5σ (4σ – 1.5σ = 2.5σ) resulting in each output having greater variability from one another. In order to account for variation, a process spread of 6σ was suggested to preserve the process under specifications even if a shift of 1.5σ were to occur (6σ – 1.5σ = 4.5σ). By establishing a standard of 6σ from the outset, the process is still highly standardized even if it shifts, thus leaving the process at 4.5σ (Swinney, n.d.). The exact reason why a shift of 1.5σ was assumed and how such a value was arrived upon has been a topic of contention. Some have argued that in a properly monitored process, such a shift would have been detected immediately and controlled. The accuracy of the 3.4 ppm figure (see Table 1) assuming a 1.5σ shift has also been under scrutiny (Voelkel, 2004). Nevertheless, Motorola's assessment of 1.5σ stood and has been considered the baseline 'standard' approximation with Six Sigma charts reporting values with this shift in mind. The bottom line is that Motorola acknowledged the existence of some form of variation, whether it is .5σ or 1.5σ, regardless of how controlled a process might be. Table 1 illustrates the percentage of acceptable values and its defective rates with and without shift according to sigma levels.

Table 1

Six Sigma Process With Shift and Without Shift (Adapted from Park, 2003).

Sigma Process Level	**With Shift of 1.5σ**		**Without Shift**	
	Acceptance Rate (%)	Defective Rate (ppm)	Acceptance Rate (%)	Defective Rate (ppm)
1σ	30.23280	697,672	68.26894	317,311
2σ	69.12300	308,770	95.44998	45,500
3σ	93.31890	66,811	99.73002	2,700
4σ	99.37900	6,210	99.99366	63.4
5σ	99.97674	233	99.99994	0.57
6σ	99.99966	3.4	99.9999998	0.002

As an example of a process shift, if an item has an original specification of 50mm, with a tolerance of +/- 2mm, and if the process goal is 3σ, the item is free to vary +/- 0.6mm (52mm - 48mm / 6σ = 0.6mm) up to a limit of 51.8mm or 48.2mm before it approaches and exceeds the threshold of being considered 'defective'. Further, if the process drifts by say, 1σ, it will degrade to a 2σ level (52mm - 48mm / 4σ = 1mm) resulting in a less standardized overall process distribution with an increased number of defectives (Figure 2).

In order to proactively avoid the negative consequences of a process shift, one would attempt to establish a process of 6σ (52mm - 48mm / 12σ = 0.3mm), resulting in a leptokurtic (low variation) distribution (Figure 3). Thus, even if the process were to drift slightly, the overall process is still highly standardized, increasing the probability that the number of defectives remain minimal. To reiterate, the purpose of Six Sigma is to attain a high process quality via standardization through the minimization of variation. Further, the most successful approach has been where one shifts away from reactively fixing something once a defective is identified, to proactively identifying and controlling causes of variation, which in turn results in a lower rate of defectives. By achieving such a standard, productivity and customer satisfaction is increased and so is profitability in some cases (Velocci, 1998).

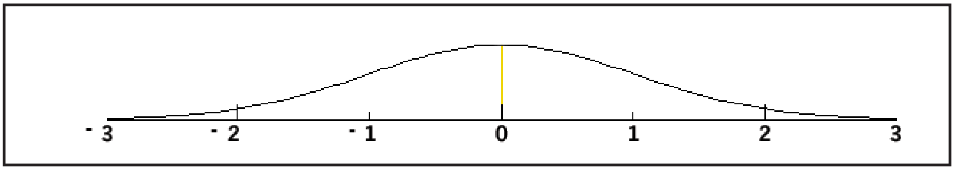


Figure 2. High variability distribution with several scores away from the target specification.

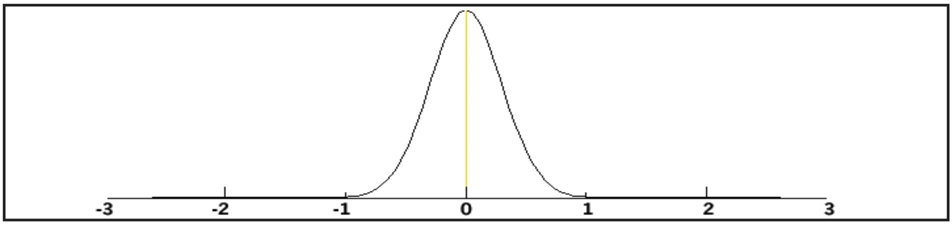


Figure 3. Low variability distribution, scores are closer to specification.

Six Sigma Management System

Motorola management considered Six Sigma a paradigm shift in the way the company operated at all levels. By involving management in the new quality thought process, a top down approach becomes possible, where all employees are trained and educated in the concept of quality and the need for the identification of causes of variation and controlling those causes (Motorola, 2004). By having all levels involved, emphasis is placed on teamwork – where multiple teams throughout the company via their respective team leaders have a singular goal of satisfying the requirements of all respective stakeholders who are recipients of whatever process output. These processes include anything from payroll to document processing, shipping, and even marketing. The result was that Six Sigma has evolved from an operationally focused metric into a management system.

Although process standardization is the goal, Six Sigma is distinctive in that it provides a structure in which to attain reduction in variation through the process improvement methods. DMAIC (Define, Measure, Analyze, Improve, and Control) is the typical tool used for making such improvements.

DMAIC

DMAIC has been defined as a ‘rigorous, structured, and disciplined’ approach to process improvement (Rath & Strong, 2003). The tools contained within DMAIC are simple but effective, and have been available in one form or another in several previous quality methodologies such as Total Quality Management.

According to Park (2003), Six Sigma is simply an evolution of Total Quality Management (TQM), which in turn was built upon Total Quality Control (TQC), Statistical Quality Control (SQC), and Quality Control (QC). Whereas TQM and earlier quality methodologies provided a multitude of statistical tools aimed at achieving and maintaining a high level of quality, Six Sigma has provided a structured framework in which these tools may be more effectively used. Figure 4 lists some of the most commonly utilized tools in each phase.

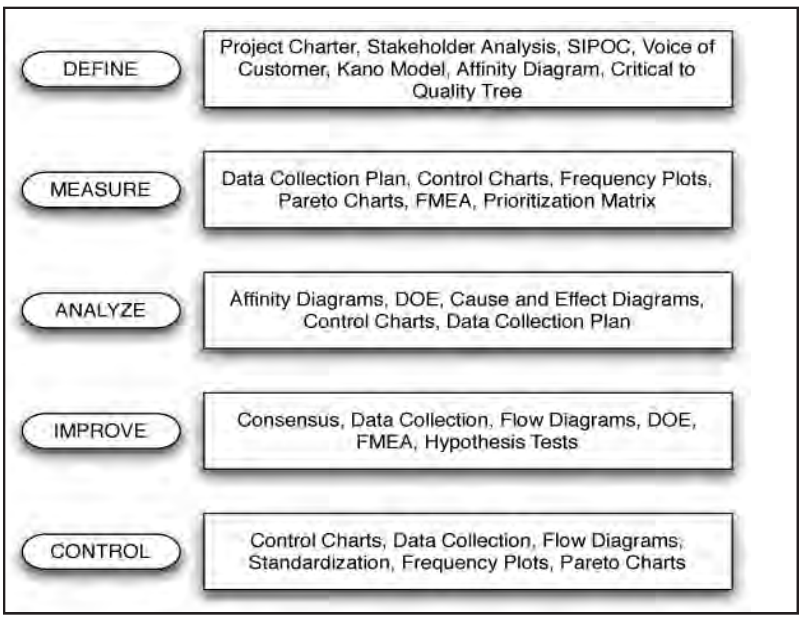


Figure 4. Typical quality tools employed throughout the DMAIC process (Adapted from Rath & Strong, 2003).

The following is a general overview of the major objectives in each step of the framework provided by Six Sigma:

Define. This phase involves a systems engineering approach, where the purpose and scope are defined together with background and historical information. Study goals are defined and so are limitations as to how far the study is to go and what it can bring to the overall operation. A stakeholder analysis is also performed, where each of those involved (in an airline setting this may include managers, analysts, pilots, maintenance) defines what such a study is expected to accomplish for them.

Measure. The priority in this step is placed on the improvement effort. Historical information and other data relevant to the subject at hand are gathered. Using this information, the source of the problem is identified for further analyses. The current process sigma is also determined at this point.

Analyze. Based on the current process performance and knowledge of the source of the problem as determined in the previous phase, the focus then shifts towards identifying root causes. Root causes can range from poor communication between departments, lack of resources allocated to the wrong places, and even the wrong data being collected. Techniques such as Design of Experiments (DOE) could potentially identify variables that were initially unforeseen. Data mining is yet another breakthrough technique in the quest to identify causal factors and trends among a multitude of data.

Improve. In this step, candidate solutions are introduced and implemented. The purpose is to verify that the proposed improvements solve the issue at hand. Some issues might be resolved completely without further intervention; however, others require even deeper understanding. In some cases, a lot of data is present and experiments can be done to determine the complexity of the issue.

Control. Suggested solutions in the previous step are prepped for implementation. The focus of this step is standardization, which will ultimately result in a decrease in variation and, thus, a higher sigma process level.

The application of the DMAIC framework has been successful across many industries, regardless of the processes involved. And though they do not use these specific terms themselves, the FOQA Rule (14 CFR, Part 13) and the associated guidance provided to FAA Inspectors responsible for oversight of FOQA programs (HBAT 00-11) both indicate a requirement that mature FOQA programs possess the attributes of continuous improvement (that DMAIC inherently supports). It is in the Measure and Analyze steps of DMAIC that Six Sigma techniques offer the greatest power.

The parameters recorded during flight allow for a FOQA air carrier to monitor adherence to standard flight protocols. Each parameter can be monitored for variance based on set tolerance thresholds as determined by the air carrier upon appropriate validation. For example, a target value of 165 knots could be established for a certain phase of flight, with a maximum allowable variation of ±10 knots. Any exceedance (which in Six Sigma terms can be considered a 'defect') of these limits is flagged as an 'event', which is differentiated by severity levels. Therefore, a recorded parameter of 172 knots might be considered a level 1 severity event, while an exceedance of 180 knots could be considered a severity 3.

When excessive numbers of severity 1 and 2 events are detected by the FDM software, airline managers might elect to re-evaluate the tolerances since they might be too strict. However, when a severity 3 is detected, it usually points to a potentially dangerous violation of standard procedures; thus, they usually warrant close examination. If an airline continues to detect excessive numbers of severity 3 or other events after adjusting severity thresholds, the potential for an incident or accident may be indicated.

FOQA's proactive nature means that it functions by concentrating on level 1 and 2 events, proactively implementing remedial action and standardizing the operations in order to avoid level 3 events from occurring. In the U.S., commercial air transportation is already highly standardized and level 3 events are rare, but they do occur. Examples of level 3 events are tail strikes during takeoff, and overshooting or undershooting runways during final approach due to energy mismanagement. The rarity of these events makes it problematic to utilize rate-based methods that depend on events that have already occurred in order to estimate the chances of any future occurrences.

To illustrate, for an air carrier operating thousands of flights per month, FOQA trend data will be increasingly abundant with commonly occurring events such as speed or pitch violations. As data is collected and analyzed, the distribution will eventually become normalized, allowing for proper predictive statistics. However, for extremely rare events such as tail strikes, the distribution will not likely be normal, but rather highly skewed due to the extended amount of time without any occurrence. There will not be enough data to support proper predictive statistics.

Tail Strikes

Tail strikes are serious events with historically low rates of occurrence. Some tail strikes are so severe that they are declared accidents due to the extensive damage to the aircraft. These can prove costly in many ways, such as in maintenance costs and damage to an air carrier's reputation.

As a hypothetical example of how Six Sigma techniques could aid a FOQA study, a newly formed air carrier is interested in the topic of tail strikes during takeoff and what Six Sigma techniques can offer. Assuming the carrier has been operating at a rate of 500 flights per month, the flight safety department would like to determine the probability of a tail strike occurring during takeoff based on a year's worth of data gathered via the FOQA program. The aircraft manufacturer established that a tail strike could occur if the takeoff angle reaches a certain critical angle with the main undercarriage oleo fully compressed. The air carrier established as a standard procedure a rate of rotation after takeoff of 2 to 3 degrees per second to a pitch attitude of 15 degrees.

Historically, tail strikes during takeoff have involved several different variables. Some of the most commonly attributed causal factors are:

1. Improperly trimmed stabilizer
2. Improper rotation speed
3. Improper flight director use
4. Excessive rate of rotation

No tail strikes have yet occurred, and the air carrier would like to minimize as much as possible the chances of one happening. Utilizing Six Sigma's DMAIC methodology, the air carrier would like to determine what its current process level is and what the probabilities are of a tail strike during takeoff given its current process capability. This study is presented below according to the DMAIC structure.

Define. During the define phase, the underlying motivation was to idetify methods in avoiding any embarrassing and costly events from occurring. Tail strikes during takeoff are the FOQA topic selected and the decision was made to focus on one aircraft. The objective is to determine the current process level and the potential for future tail strikes.

Measure. The aircraft is fully FOQA equipped. Based on relevant parameter data, it was determined that the mean for the parameter 'Max Takeoff Pitch' of all flights up to this point was 15.5 degrees and the standard deviation was 1.67, thus establishing the process sigma at 3.89 with an exceedance rate of .005% - equivalent to a potential of one tail strike every 19,951 flights with ~39 months between each occurrence.

Analyze. The calculations indicate that the air carrier is due to experience a tail strike in about two more years of operation if no change is made. Therefore, several different scenarios are considered. For example, if the mean (average max takeoff pitch) of 15.5 degrees is maintained, but the standard deviation is decreased to 1.5, the exceedance rate would improve to .0007% (equivalent to a process sigma of ~4.33). Thus, approximately one tail strike every 134,127 flights is expected, equivalent to ~268 months before the event is due to occur.

Another scenario would be if the mean were decreased to 14.5 and the standard deviation maintained at 1.67. This would result in a process sigma of 4.49, where the exceedance rate of .0004% would be equal to approximately one tail strike every 280,817 flights, equivalent to ~561 months before one is due to occur. Hence, it is clear that even slight improvements in standardization significantly decrease the probability of a tail strike occurrence. Additionally, if the standard deviation remained the same, but the mean of the scores improved, significant reductions in the probability of a tail strike occurrence is also possible.

Naturally, one should not adopt a false sense of security by depending solely on these predictive rates, as the nature of probability theory dictates that the events can occur more or less frequently than expected. However, since probability is based on what is likely to occur, a prudent airline will try to get the odds on its side. Finally, this approach is only one of several factors that have a bearing in determining the likelihood of a tail strike. There have also been efforts by aircraft manufacturers such as Boeing's implementation of the 'tail strike protection application' within the flight control system software of the B777-200LR and 300ER variants (Louthain, 2005). This demonstrates the current interest in every sector within the aviation industry in flight safety.

Improve. Given the analyses of possible scenarios, stakeholders are presented with various solutions. These may include forming an informational campaign for the pilots demonstrating that even slight improvements in standardization and adherence to flight procedures can significantly decrease the likelihood of a serious event occurring. Another choice would be modifying current standard flight procedures to reduce the pitch attitude from 15 degrees to 14 degrees and, thus, significantly reducing the chances of a tail strike occurring (even if the standard deviation remained constant).

It is worth noting that given the complexity of flight operations, the possibility of creating unintended consequences is an important factor to keep in mind when exploring improvement strategies. For example, the reduction of initial rotation pitch attitude described above might result in compromised obstacle clearance or noise abatement. As with any intervention strategy, a full consideration of the consequences is necessary before proceeding with the plan. Once having defined the potential effects of the intervention, wise use of FOQA can give valuable information on all of those effects, as the DMAIC process proceeds from Improvement to Control.

Control. Whichever solution is chosen, relevant data can be continuously monitored to verify the effectiveness of the changes undertaken utilizing tools such as process control charts. This hypothetical case study is only one of several possible studies an air carrier could perform with an existing FOQA program by adopting Six Sigma techniques. Advanced methods such as data mining and design of experiments (DOE) could also provide a deeper insight into tail strikes. For example, what is the relationship between energy and tail strikes? Also, are there any other monitored aircraft parameters that might have potential influence in a tail strike occurrence? Future possibilities also include data mining the Aviation Safety Action Program (ASAP) database and correlating the information with FOQA databases.

Discussion

Flight Operations Quality Assurance has been one of the most highly regarded and potentially effective airline safety initiatives to emerge in the past 20 years. It is a program based on quantifiable, objective data collected from the air carrier aircraft's data recording system. On some modern aircraft, over 2000 parameters each second are recorded. The FOQA system uses expert software to analyze the data from individual flights of interest, or aggregated data from multiple flights in order to examine trends that may affect safety. Unfortunately, with very few exceptions, the analysis of FOQA data has been limited to relatively simple statistical methods. It has been surmised that the application of more sophisticated quality and statistical methods may increase the effectiveness of the program and the air carrier's return on investment (Stolzer & Halford, 2004).

Six Sigma is a structured, data-driven approach to eliminating defects. The primary objective of the Six Sigma methodology is the implementation of a data-based strategy that focuses on variation reduction and process improvement through the application of Six Sigma improvement projects. DMAIC – Define, Measure, Analyze, Improve, and Control – is the method used to engage in process improvement. It was asserted that Six Sigma methods might be effectively used in FOQA programs, especially for addressing very infrequently occurring events.

An exemplar case study was presented using Six Sigma's DMAIC methodology on a safety problem, i.e., tail strikes during takeoff. The process sigma was calculated to be 3.89 with an exceedance rate of .005%, which equates to a potential for one tail strike every 19,951 flights with ~39 months between each occurrence. The effect on the process sigma of varying the mean and standard deviation was explored. Stakeholders were presented with various solutions to decrease the probability of a tail strike from occurring.

A disciplined quality approach to improving safety is needed in the airline industry. Airlines would benefit by increasingly embracing and employing quality principles in designing, implementing, and managing safety programs, including FOQA. Six Sigma is one quality-based program that may be used to increase the effectiveness of FOQA, particularly for process improvement initiatives. Whether an airline employs Six Sigma or various other methods in its safety improvement efforts, quality in airline safety must be the goal.

References

Global Aviation Information Network (2001, December). *Survey of analytical processes and requirements for airline flight safety management.* Chevy Chase, MD. Retrieved November 2005 from http://204.108.6.79/products/documents/Survey_Analytical_Processes.pdf

Harry, M. (n.d.). *Where did the name "Six Sigma come from?* Retrieved November 2005 from http://www.isixsigma.com/forum/ask_dr_harry.asp?ToDo=view&questId=109&catId=10

Louthain, S. (2005, October 21). *Flight test journal: Protecting your tail.* Retrieved November 2005 from http://www.boeing.com/commercial/777family/200LR/flight_test/archives/2005/10/index.html

Motorola, (2004). *Motorola university: Free six sigma lessons.* Retrieved November 2005 from http://www.motorola.com/content/0,,3069-5787,00.html#

National Aeronautics and Space Administration (NASA). (2004). *Coast to coast simulation tests new air traffic management concepts.* Retrieved November 2005 from http://www.nasa.gov/vision/earth/improvingflight/DAG-TM.html

NTSB. (2004). Safety Accomplishments. In *NTSB Accomplishments, 1993-2000*. Retrieved July 2004, from http://www.ntsb.gov/Recs/accomplishments_93_00.htm

NTSB. (2005). *Aviation Accident Statistics.* Retrieved November 2005 from http://www.ntsb.gov/aviation/Table6.htm

Park, S. H. (2003). *Six Sigma for quality and productivity promotion.* Retrieved November 2005 from http://www.apo-tokyo.org/00e-books/IS-09_SixSigma.htm

Phillips, E. (2002). Recording sessions. *Aviation Week & Space Technology, 157*(18), 19.

Rath & Strong Management Consultants. (2003). *Six sigma pocket guide*. Lexington, MA: Rath & Strong Management Consultants.

Stolzer, A. & Halford, C. (2004). Wanted: quality in airline safety. Journal on *Aviation/Aerospace Education and Research (JAAER), 13*(3), 27-32.

Swinney, Z. (n.d.) *1.5 sigma process shift explanation.* Retrieved November 2005 from http://www.isixsigma.com/library/content/c010701a.asp

Velocci, A. (1998). Pursuit of Six Sigma emerges as industry trend. *Aviation Week and Space Technology, 149*(20), 52-53.

Voelkel, J. (2004). *What is 3.4 parts per million?* (Technical Report 2004–2) [Electronic version], Rochester, NY: Rochester Institute of Technology, Center for Quality and Applied Statistics. Retrieved November 2005 from http://www.rit.edu/~636www/about/TR2004-2.pdf

Weener, E. F. (1990, December). *Control of crew-caused accidents.* Paper presented at Third Annual Society of Automotive Engineers Symposium on Human Error, Dallas, TX.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Comparison of Four Quantitative Methods of Measuring Human Judgment Performance in a Simulated Horizontal Air Traffic Conflict Prediction Task

Ellen J. Bass
University of Virginia
Department of Systems and Information Engineering
P.O. Box 400747
151 Engineer's Way
Charlottesville, VA 22904

Abstract

An experiment using a simulated air traffic conflict prediction task with sensor noise was conducted. Measures based on four quantitative methods for modeling human judgment performance in uncertain environments (signal detection theory (SDT), fuzzy signal detection theory, lens modeling, and skill scores) were compared. In addition to individual differences, gender, noise level, and session (experience) effects and their two-way interactions were investigated. Both a fuzzy signal detection theory sensitivity measure and the lens model achievement indicated a significant gender effect, with males outperforming females. Only the skill score indicated a significant gender-noise level interaction and only the fuzzy signal detection theory bias measure indicated a significant noise level-session interaction. The implications of this effort are that analysts should not rely on the results from any single methodology when analyzing human judgment performance in uncertain environments.

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.

Introduction

Judgment is a critical component of many human activities. People have to make judgments every day in situations where the environment is changing and information is ambiguous or not current. Measuring human judgment in such environments is complicated by the multiple loci of uncertainty in the judgment process: in the judge, in the environment, and in the relationship between the judge and the environment.

Quantitative measures can aid in the evaluation of human judgment performance in dynamic, uncertain environments. However, most studies only consider one method for modeling performance and therefore analysts have little information on which method to use in different situations. This paper compares four quantitative methods of measuring human judgment performance for the same noisy experimental judgment task. The focus is not the specific results of a particular measure for the artificial task used in this study but rather on the similarities and differences between the measures derived from the methods. The three existing quantitative methodologies that consider the three loci of uncertainty in the judgment process are included: signal detection theory (SDT) (Green & Swets, 1989), fuzzy SDT (Parasuraman, Masalonis, & Hancock, 2000), and lens modeling analysis (Cooksey, 1996). In addition, the skill score (Murphy, 1988) is included as it has been combined with lens modeling analysis to extend the latter's diagnostic capabilities (Stewart, 1990). In this article, the four methodologies are first described and the purpose of the research detailed. Second, the experimental task used in this research is described. Third, the methods used to conduct the empirical study are discussed. Fourth are the results. The paper concludes with a discussion of the results and future work.

Signal Detection Theory

Traditional SDT was developed to model the detection of an event in a noisy environment (Green & Swets, 1989). It focuses on the detection process in the presence of an evidence variable, "X," and noise. SDT assumes that the judge has a cutoff value, C_h. When the properties of X exceed C_h, the judge would assert that the signal is present. The combinations of the states of the world (signal with noise present as opposed to noise only) and the two possible responses ("yes," there is a signal or "no," there is no signal) create four classes of joint events. Two are correct responses: hit (responding "yes" when the signal is present) and correct rejection (responding "no" when it is a noise only event) while two are errors: false alarm (responding "yes" when it is a noise only event) and miss (responding "no" when the signal is present) (Table 1). From the four possibilities, four probabilities are calculable:

- P(H): Probability of hit (number of hits/number of signal with noise present events)
- P(FA): Probability of false alarm (number of false alarms/number of noise only events)
- P(M): Probability of miss (number of misses/number of signal with noise present events)
- P(CR): Probability of correction rejection (number of correct rejections/number of noise only events)

Table 1
Signal Detection Theory Outcomes

Response

	Signal + Noise	Noise only
Yes	Hit	False alarm (Type I error)
No	Miss (Type II error)	Correct rejection

SDT uses two parameters to model detection. Sensitivity is an index of the judge's ability to distinguish the signal with noise present from noise only events. Response bias is the judge's tendency to respond positively or negatively as a function of the four outcomes and the likelihood of a signal being present. With the assumptions of normality and of equal variance for the signal and noise distributions, sensitivity is the distance between the means of the signal and the noise scaled to the standard deviation of the noise distribution. The response bias is the likelihood ratio that an effect of the cutoff criterion is due to signal plus noise as opposed to noise alone.

If the normality assumptions are violated, non-parametric sensitivity and bias measures can be computed using P(H) and P(FA) (Green & Swets, 1989):

$$d(A) = 0.5 * (P(H)+(1-P(FA))) \qquad \textit{Equation 1}$$

$$C = -0.5 * (Z(P(H)) + Z(P(FA))) \qquad \textit{Equation 2}$$

The value for sensitivity, d(A), falls between zero (the worst possible performance when P(H) is 0 and P(FA) is 1) and one (the best possible performance when P(H) is 1 and P(FA) is 0). Sensitivity is 0.5 when P(H) and P(FA) are equal. With respect to the bias measure, C, a negative bias indicates a liberal response (i.e., where liberal denotes a tendency to respond "yes" meaning that there will be a high false alarm rate compared to the miss rate) while a positive value indicates a more conservative stance.

Fuzzy Signal Detection Theory

If the signal and/or the response is continuous rather than Boolean, fuzzy logic extensions to SDT should be considered. Fuzzy set theory developed from the need to model approximate reasoning with imprecise propositions (Zadeh, 1965). Fuzzy logic has been used to model judgment in several domains such as medicine where the measurements provided by a laboratory test can be of limited precision and where the border between normal and pathological is not exactly clear (c.f. Sanchez, 1979). Fuzzy logic approaches have recently been used to model human-automation interaction to guide design (Parasuraman, Masalonis, & Hancock, 2000) and training (Campbell, Buff, & Bolton, 2000).

Fuzzy set theory extends traditional set theory by allowing the assignment of membership functions to items in a set as opposed to only two values (namely 0 and 1). A value is assigned to each element of the universal set signifying its degree of membership in a particular set with unsharp boundaries. Formally, if X is the universal set, then the membership function μ_A defines fuzzy set A as follows:

μ_A: X->[0,1]   *Equation 3*

For example, the membership function for the fuzzy set "horizontal air traffic conflict" could assign membership grades based on the horizontal separation distance between the aircraft and for the fuzzy set "vertical air traffic conflict" based on vertical separation distance.

Fuzzy measures provide ways to indicate the degree of evidence or certainty of an element's membership in a crisp set. With signal detection, where there are two non-overlapping categories (signal and noise), fuzzy measures can be used to assign the degree of set membership to both sets. Parasuraman, Masalonis, and Hancock (2000) introduced the four steps required for fuzzy SDT analysis described next.

Select mapping functions. The signal mapping function maps variables describing the state of the world into the set S (signal) with some membership degree in the range [0,1]. Mapping functions can map a single variable into the range [0,1], or can operate on some combination of variables. Similarly, the response mapping function assigns the result into the set R (response) with some membership degree in the range [0,1] based on a judgment of confidence that the signal is present, and/or the signal's perceived or reported severity, strength, or criticality.

Implication functions. The observed values of S and R are used to derive fuzzy set memberships in the SDT outcomes of H, M, FA, and CR. Parasuraman, Masalonis, and Hancock (2000) proposed the following implication functions for this purpose:

H = min (S, R)   *Equation 4*
M = max (S-R, 0)   *Equation 5*
FA = max (R-S, 0)   *Equation 6*
CR = min (1-S, 1-R)   *Equation 7*

Calculate hit and false alarm rates. The hit rate (HR) is calculated by dividing the sum of the hit memberships of each event across the trials, by the sum of the signal membership values (S). To calculate FA rate (FAR), the sum of the FA memberships of all events is divided by the sum of the not-signal membership values (1-S).

Computation of fuzzy SDT measures. The final step involves the computation of measures from the fuzzy hit and false alarm probabilities. This step is essentially the same as those computed for SDT.

Lens Modeling

The lens model is based on probabilistic functionalism, which designates the organism-environment interaction as the primary unit of study (Brunswik, 1956). It considers both internal (cognitive) and external (environmental) aspects of judgment. Making judgments includes acquiring environmental information and transforming that information into an assessment of the environmental criterion. To make a judgment, a judge considers one or more cues (i.e., pieces of information from the environment). The cues must be interpreted and integrated to form a judgment. For some judgment tasks, this process must be repeated over time to discern trends. The cues may have associated uncertainty, as they may be incomplete, noisy, or otherwise erroneous. Uncertainty has implications for the judgment process as it limits both the predictability of the environmental criterion and the judgment. When investigating the judgment process, one must therefore consider the information available, the relationship between the environmental criterion and that information, and the relationship between the information and the judge (Brunswik, 1956). That is, investigating judgment achievement (i.e., how well a judge's judgments correspond to the environmental criterion) requires the consideration of both the environmental criterion-information and the information-judge relationships (Brunswik, 1956).

Brunswik (1956) conceptualized the goodness of the environmental criterion-information relationship as a degree of correspondence called ecological validity (i.e., the utility of the information available in the environment in determining the environmental criterion) and the goodness of the information-judge relationship as a degree of correspondence called cue utilization (i.e., the judge's utilization of the available information in making a judgment).

The lens model depicted in Figure 1 illustrates how uncertain information can affect judgment achievement with respect to ecological validity and cue utilization. The left side of Figure 1 depicts the task environment in terms of the cues available (the X_i's in the figure) and the environmental criterion (Y_e in the figure) to be judged. Cues and the criterion are related by statistical correlations known as ecological validities (the r_{ei}'s in the figure). If the specific cues available in the environment reflect the true state of the environment, then those cues have high ecological validity. The right side depicts the judge using the cues to render a judgment about the environmental criterion. Correlations between the cue values and the judgments (Y_s in the figure) are the cue utilizations (the r_{si}'s in the figure). If a judge uses specific cues in making a judgment, then those cues have high cue utilization. The particular pattern of cue utilizations exhibited by a judge determines the cognitive judgment strategy. Achievement (the top arc reflecting the correlation between the judgments and the environmental criterion) will be maximized when the pattern of cue utilizations mimics the pattern of ecological validities. In general, judgment achievement will be higher when the patterns of cue utilization match those of the ecological validity.

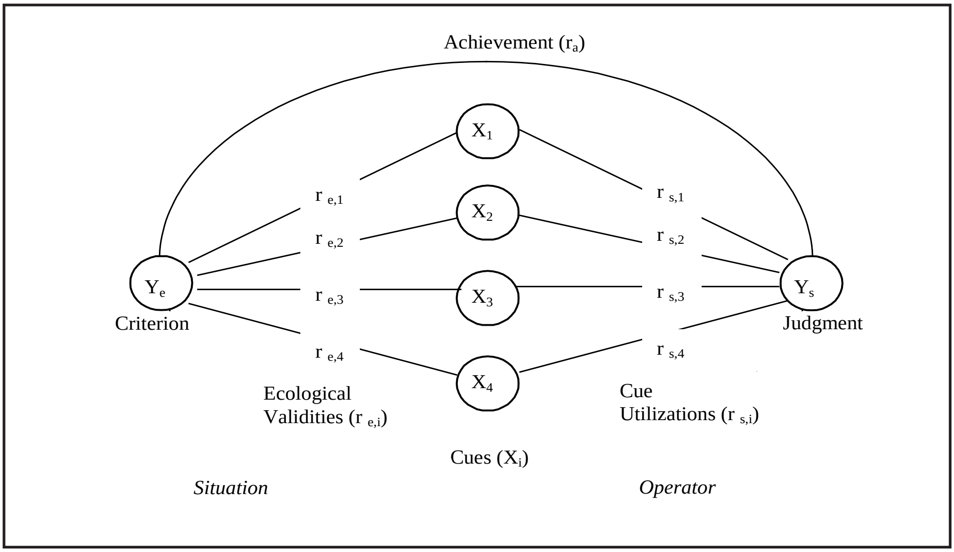


Figure 1. Double System Lens Model of Judgment

The lens model structure yields the lens model equation (LME) (Hursch, Hammond, & Hursch, 1964; Tucker, 1964):

$$r_a = G \, R_e \, R_s + C\sqrt{1-R_e^2}\,\sqrt{1-R_s^2} \qquad \textit{Equation 8}$$

where:
r_a = Achievement
G = Linear Knowledge
R_e = Environmental Predictability
R_s = Cognitive Control
C = Nonlinear Knowledge

With the LME, the lens model aids in understanding the source of less than perfect judgment by decomposing achievement. The first term in the LME is the product of linear knowledge, environmental predictability, and cognitive control. R_e is calculated as the multiple correlation of the environmental linear regression model. It represents a limit on judgment performance based on the predictability of the task environment. G indicates the level of judgment performance if the environment and the judge were completely linearly predictable. It is calculated as the correlation between the predictions of the two (environmental and cognitive) regression models. The adequacy of a judgment strategy (in terms of beta weights in the cognitive regression model) represents the linear knowledge. The consistency with which a judge can execute his or her strategy is captured by cognitive control, R_s, calculated as the multiple correlation from regressing the judgments on the cue values.

The second term in the LME deals with nonlinear effects not captured by the purely linear model represented in the first term. The values of R_e and R_s appearing in this term have already been discussed. Nonlinear knowledge is calculated as the correlation between the residuals of the environmental and the cognitive linear regression models. Its role is to identify if the judge is capturing non-linear components in the environment that are not captured in a linear model.

Skill Score

As an alternative to regression-based approaches, Mean Square Error (MSE) has been used to measure judgment performance (Murphy, 1988) where n is the number of judgments:

$$MSE_Y = \frac{1}{n}\sum_{i=1}^{n}(Y_{si} - Y_{ei})^2 \qquad \textit{Equation 9}$$

Different decompositions of MSE have been considered by researchers (c.f., Cooksey, 1996; Murphy, 1988). The form of the judgment performance standard is one such difference. Stewart (1990) uses a constant judgment based on the average value of the criterion. The correspondence (i.e., "goodness") of the standard is defined as MSE_R where $\overline{Y}e$ is the mean of the criterion:

$$MSE_R = \frac{1}{n}\sum_{i=1}^{n}(Y_{ei} - \overline{Y}_{ei})^2 \qquad \textit{Equation 10}$$

The skill score (SS), defined as the ratio between the MSE of the judgments and the MSE of the standard subtracted from unity, is a derived measure of judgment performance (Murphy, 1988). SS is positive when the judgments are better than the standard (MSEY < MSER). When the SS is zero, the judgments are as good as the standard (MSER =MSEY). When it is negative, the judgments are worse than the standard (MSEY > MSER).

$$SS = 1 - \left[MSE_Y \big/ MSE_R \right] \qquad \textit{Equation 11}$$

Murphy (1988) developed the SS to decompose the MSE by substituting the equations for MSEY and MSER into the form of the SS above:

$$SS = (r_a)^2 - \left[r_a - \left(\frac{s_{Ys}}{s_{Ye}} \right) \right]^2 - \left[\frac{\overline{Y}_s - \overline{Y}_e}{s_{Ye}} \right]^2 \qquad \textit{Equation 12}$$

The first term, the square of the LME achievement, has been described. The second term, called conditional bias or regression bias, measures whether the judge has appropriately scaled judgmental variability to situational variability. Conditional bias illustrates a tendency to produce judgments on an interval different from that found in the true situation. The third term, unconditional bias or base rate bias, measures the overall bias in the judgments, illustrating a tendency to over- or under-estimate the criterion. This bias equals zero only when the mean of the judgments equals the objective base rate.

This MSE-based approach provides the advantage that it can be partitioned into three distinct components representing shape, scale, and magnitude. It also makes no commitment to the cues used by the judge. Its disadvantage is that it penalizes larger errors compared to smaller errors.

Purpose of This Research

To compare the four methods, data from the same experiment were analyzed with each. The task (described in detail in the next section) is to judge the probability that a simulated aircraft will conflict with another simulated aircraft. To provide a basis for analysis, independent variables included the level of uncertainty (i.e., sensor noise), gender, and experience with the task (determined by the experimental session). The dependent variables appear in Table 2.

Table 2
Dependent Variables

Description	Methodology
Sensitivity: d(A)	SDT
Bias: C	SDT
Fuzzy sensitivity: F d(A)	Fuzzy SDT
Fuzzy bias: F C	Fuzzy SDT
Achievement: ra	Lens Model
Linear Knowledge: G	Lens Model
Cognitive Control: Rs	Lens Model
Nonlinear Knowledge: C	Lens Model
Skill score: SS	Skill Score
Conditional bias: CB	Skill Score
Unconditional bias: UB	Skill Score

Regardless of the modeling methodology employed, our hypotheses included the following:

- In general, males should outperform the females on performance measures on this spatial task (c.f., Linn, & Petersen, 1985; Maccoby & Jacklin, 1974) although dissenting opinions exist with regard to this gender difference (Caplan, MacPherson, & Tobin, 1985).
- Participants should perform better in low noise as compared to high noise conditions as it is easier to perform a task with less uncertainty (Wickens, Gordon, & Liu, 1998).
- As the participants have no previous experience with the task, they should perform better in later sessions as practice enhances performance (Wickens, Gordon, & Liu, 1998).

We expected that the traditional and fuzzy SDT sensitivity measures, the human judgment related LME measures, and the SS should reflect these patterns. Because the traditional SDT C and fuzzy SDT bias measures reflect the liberal or conservative stance of a participant (as opposed to performance), we have no a priori hypotheses with respect to these measures concerning gender, noise level, or session. However, because the aircraft collisions are unacceptable, our hypothesis is that in general the participants would be liberal and would accept more false alarms but not misses.

With respect to the four methods, our hypothesis is that traditional SDT would be sensitive to systematic patterns in the data that would not be reflected with the other methods. Depending on the cutoff used to separate “Yes” responses from “No” responses, traditional SDT can provide much different performance values as compared to the other two methods. Because traditional SDT forces probabilistic data into only two categories, judgments close to the environmental criterion but on the “wrong side” of the cutoff receive no credit. With the other two methods, there are provisions for credit being given for judgments being close to the cutoff.

An example can shed light on this concept (Table 3). Assume there are six discrete, but non-binary, environmental criteria (S). From them, six derived traditional SDT criterion values (SDT S) can be calculated using a cutoff value (such as 0.5 in Table 3). Assume that two judges provide responses, R1 and R2, which are very close to each other (R1 and R2 are only 0.02 apart on each probability judgment). Although the judges have judgments very close to each other, using the same 0.5 as the cutoff values, the traditional SDT responses (SDT R1 and SDT R2) are completely opposite (because the judgments are on the opposite sides of the cutoff). Thus the traditional SDT sensitivity values differ greatly (1 calculated using SDT R1 and 0 using SDT R2) while the fuzzy sensitivity values are close (0.727 for R1 and 0.708 for R2).

Table 3
Traditional and Fuzzy SDT Comparison Example

S	SDT S	R1	SDT R1	R2	SDT R2
0.1	0	0.49	0	0.51	1
0.2	0	0.49	0	0.51	1
0.51	1	0.51	1	0.49	0
0.6	1	0.51	1	0.49	0
0.9	1	0.51	1	0.49	0
1	1	0.51	1	0.49	0

Experimental Task

The task domain for this research is air traffic conflict prediction. A participant's task was to judge the probability that the simulated ownship would conflict with another simulated aircraft. A conflict was defined as the traffic aircraft entering the protected zone of ownship. For the experiment, the protected zone around an aircraft was a cylindrical volume five nautical miles (NMs) in radius and extending 1000 feet above and below. All aircraft maintained the same altitude and therefore only horizontal conflicts were of concern.

The models for both aircraft included uncertainty in the horizontal position based on lateral position, indicated airspeed, and course tracking errors (Table 4). For traditional SDT, the criterion was the simulated result (i.e., conflict or no conflict). For fuzzy SDT, the lens model, and the SS analyses, a continuous environmental criterion was developed. The aircraft positions were projected to a predicted point of closest approach (PCA). At the predicted PCA, the predicted horizontal miss distance ($\hat{h}$miss) was determined. The variance in horizontal position was calculated as a function of the uncertainty parameters (Table 4) and the distance and time to the PCA for each aircraft.

Table 4
Horizontal Trajectory Uncertainty

Uncertainty Parameter	Distribution
X: Lateral Position Error	Gaussian; σ = 500 meters
V: Speed Fluctuation	Gaussian; σ = 15 knots
ψ: Course Tracking Variability	Gaussian; σ = 3 degrees

The along-track error accounted for error along the longitudinal axis (that is, in the direction of ownship's track). The cross-track error accounted for error along the lateral axis (that is, to the side of ownship). These two components accounted for the total variance in the horizontal plane.

Based on $\hat{h}_{miss}$, the probability of conflict, P(C), (i.e., aircraft passing within 5 NMs) was calculated based on the estimated variance in the aircraft along-track and cross-track position. Y is an approximately normally distributed variable representing the distance between the aircraft at the PCA with position error where the mean of the point of closest approach, μ_{PCA}, is set to h_{miss}. Z is an approximately standard normal variable representing the scaled difference from the 5-mile separation distance. P(C) was determined from the cumulative distribution function of Z, φ(Z).

$$Z = \frac{5 - Y}{\sigma_{Total}} \sim N(0,1)$$
$$Y \sim N(\mu_{PCA}, \sigma^2_{Total}) \qquad \textit{Equation 13}$$
$$P(C) = \phi(Z)$$

The total variance in position, σ2Total, is approximated as follows:

$$\sigma^2_{CrossTrack} \cong 0.5 * \sigma^2_x + d^2_{own}\sigma^2_{\psi_{own}} + d^2_{intruder}\sigma^2_{\psi_{intruder}}$$
$$\sigma^2_{AlongTrack} \cong 0.5 * \sigma^2_x + 2t^2\sigma^2_V \qquad \textit{Equation 14}$$
$$\sigma^2_{Total} \cong \sigma^2_x + 2t^2\sigma^2_V + d^2_{own}\sigma^2_{\psi_{own}} + d^2_{intruder}\sigma^2_{\psi_{intruder}}$$

where

- σ_X is the standard deviation in the lateral position,
- t is the predicted time to the PCA,
- σ_V is the standard deviation of the airspeed,
- d_{own} is the predicted distance that ownship will travel from its current position to the PCA,
- σ_ψ is the standard deviation of the course error, and
- $d_{intruder}$ is the predicted distance that the intruder will travel from its current position to the PCA.

Figure 2 illustrates the environmental criterion by depicting ownship's future position at the PCA and an approximate two standard deviation position error ellipse centered on the intruder's estimated future position at the PCA. In the figure, the ownship current position is in the center and is depicted by a white airplane symbol. A compass with tick marks every 5 degrees appears on the display forty NM from ownship. An inner white-dashed range circle is located 5

NM from ownship (visualizing the horizontal component of the protected zone). Four other range circles, located at 10, 20, 30, and 50 NM, help to visualize distance from ownship. The 20 NM circle is labeled with its range. Traffic position is depicted by a triangle pointing in the direction of track heading. The ownship future position and a five NM range ring appear in gray. The traffic position error ellipse appears in green. Together the two future positions show where the algorithm predicts the traffic would be in relation to ownship at the PCA. In Figure 2, probability contours and a descriptive legend are color-coded reflecting the algorithm's probability of conflict estimates.

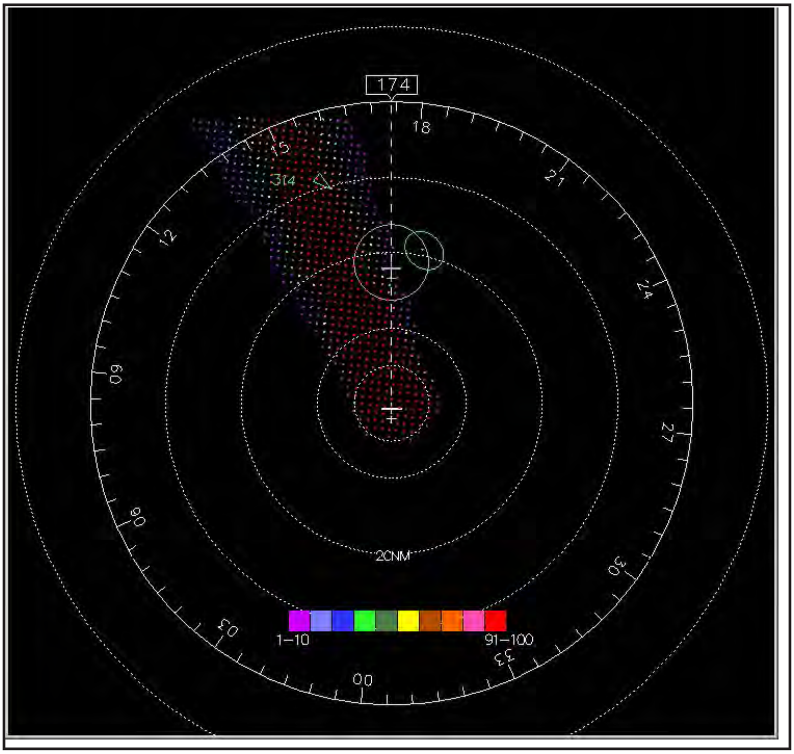


Figure 2. Visualization of the Environmental Criterion

Method

Participants

Eight male and eight female undergraduates in the same engineering program participated as paid volunteers. Their ages ranged from 21 to 25 years old (mean=22.6; variance=1.2). None had previous experience with the experimental task.

Apparatus

The experimental testbed included a low fidelity PC-based part-task aviation software suite (Pritchett & Ippolito, 2000). The testbed had three main sources of information: the Primary Flight Display (PFD), the Cockpit Display of Traffic Information (CDTI), and the data entry area. Because ownship maintained heading, airspeed, and altitude, the PFD never changed.

Traffic appeared on a CDTI in green (Figure 3). The 8-inch-by-8-inch display was “track up” (i.e., oriented based on ownship heading). A one-line data block, located behind the traffic symbol, displayed the indicated airspeed in knots. The CDTI was updated once a second. The uncertainty in the experiment was manifested by changes in the traffic position and speed at each simulation time step. The traffic would appear to jump from horizontal position to horizontal position and to change heading and speed.

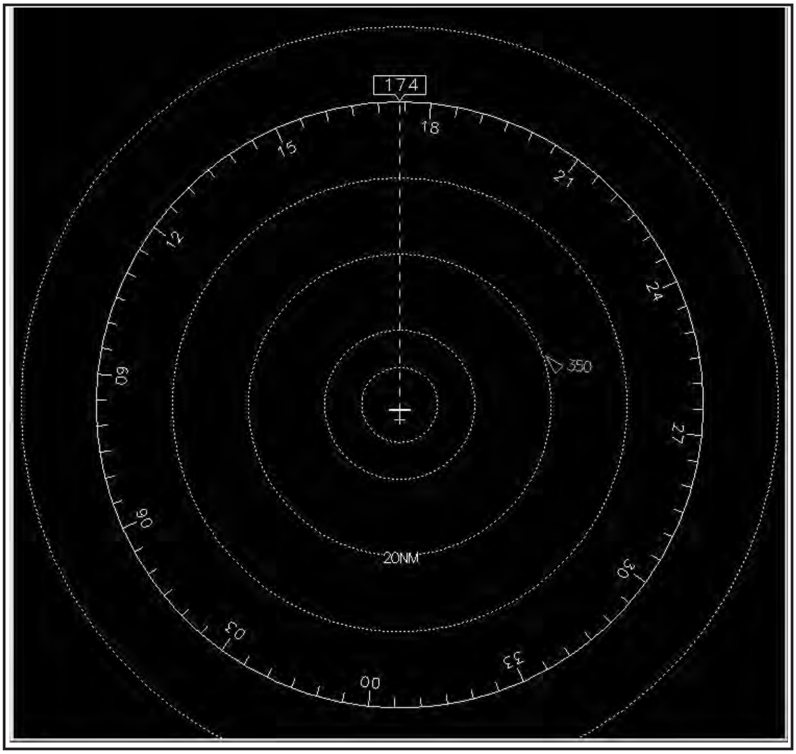


Figure 3. Cockpit Display of Traffic Information

Each trial ran for a random preview time uniformly distributed between 15 and 30 seconds before pausing for the judgment. At the judgment time, a data entry area collected the probability judgments and provided control to initiate the next trial. To enter a judgment, the participant used a slide bar or moved a gray slide bar knob. Selecting the “SUBMIT” button caused the trial to continue. The outcome of the trial (i.e., allowing the participant to monitor whether the traffic conflicted with ownship and where it was relative to ownship at the PCA) was the only feedback provided. To shorten the time for the participant to get feedback (i.e., to watch the scenario unfold after entering a judgment), the simulation update rate was increased to five times faster than real-time.

The simulator includes a scripting function that allows for the definition of experimental trials. For the experiment, 180 trials were created. Five airspeeds (300, 350, 400, 450, and 500 knots) were crossed with six headings (+/- 45^{o}, +/- 90^{o}, and +/- 135^{o}) to make 30 heading/speed combinations. Six bearings were randomly selected for each combination.

Procedure

The data analyzed herein were collected in two experimental sessions per day for the first two days of the five-day experiment (Bass, 2002). The first day started with an orientation where participants signed a consent form, were introduced to the concept of an air traffic conflict, and then viewed fast time simulations with traffic flying at different relative heading/airspeed combinations. In the experimental sessions, each trial ran for a random preview time between 15 and 30 seconds. Then each trial froze until the participant made a judgment. After the participant submitted the judgment, the trial continued in fast-time until the participant started the next trial. Each day participants judged 45 trials (one session), took a fifteen-minute break, and then judged 45 more trials (another session). The participants were debriefed at the end of the five days.

Experimental Design and Statistical Analysis

A manipulation allowed an investigation of measuring the effects of noise level. In the "high noise" condition, the uncertainty matched the horizontal position error of Table 4 while the "low noise" condition was half as noisy. Eight participants, four male and four female, were assigned to each noise level. The experimental design was nested-factorial with participants nested within gender and noise level. Each participant experienced four groups of 45 experimental trials (where the order of the trial groups was partially counter-balanced across the participants).

For the experiment, the participants and the environmental criterion provided judgments between 0 and 100%. In the traditional SDT analyses, participants' judgments below 50% were considered as "No, there is no signal present" while judgments at or above 50% were considered as "Yes, there is a signal." The environmental criterion was based on whether or not the traffic would actually conflict.

For the fuzzy SDT analyses, the participants' judgments and the environmental criterion (the probability of conflict calculation described above) were used directly. However, the tails of these distributions cut off sharply. For the LME parameter estimates and for the skill scores, the participants' and the criterion judgments were therefore transformed (Montgomery, 2000) before calculating the independent measures:

$$y = \sin^{-1}\left(\sqrt{x}\right) \qquad \text{Equation 15}$$

For the resulting correlations from the LME, the parameters were transformed for the ANOVA analyses using Fisher's r to zr transformation (as recommended in Cooksey, 1996), where r corresponds to the parameter, and zr the transformed parameter:

$$z_r = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) \qquad \text{Equation 16}$$

Linear mixed models with repeated measures ANOVA analyses were conducted where the subjects were the participants and the repeated measurements were the sessions. The fixed main and interaction effects included gender, noise level, session, and all two-way interactions. In the model, participants were nested within noise level and gender.

Results

All results are reported as significant at the 0.05 level. Trends are reported at the 0.10 level. The dependent measures were calculated using the Cognitive Systems Engineering Educational Software (CSEES) (Bolton & Bass, 2005). The statistical results were calculated using SPSS 13.0 for Windows.

SDT Sensitivity

In general, sensitivity was poor due to a relatively high false alarm rate (Table 5). No fixed main or interaction effects were significant. There was a trend for the gender-session interaction ($F_{3,20.922}=2.553$; $p=0.083$) where the males performed well right away but the females tended to get better in the last session (Figure 4).

Table 5

Summary of SDT Results

	p(H)	p(FA)	d(A)	C
Mean	0.761	0.580	0.590	-0.487
Variance	0.019	0.016	0.0063	0.106

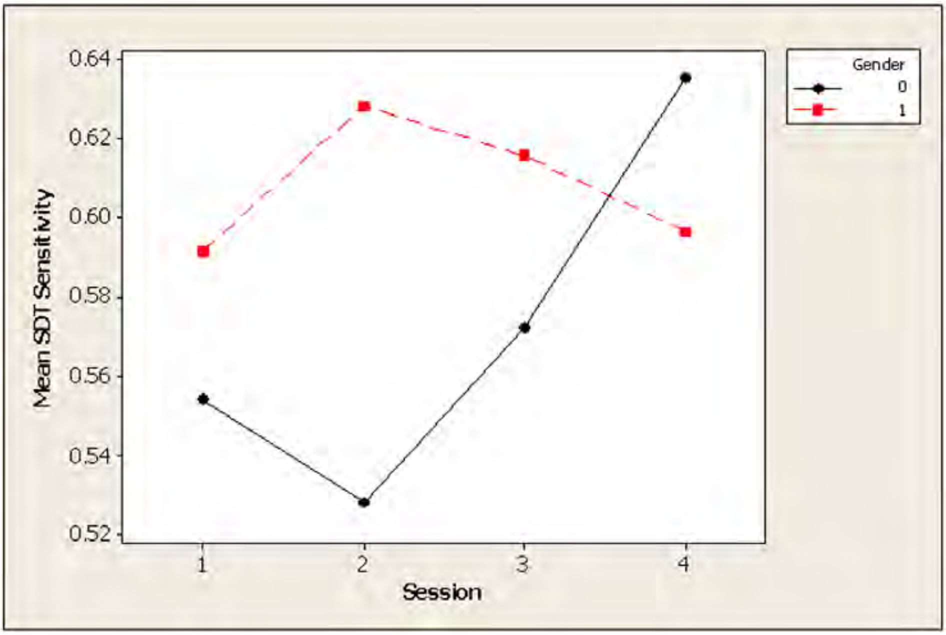


Figure 4. SDT Sensitivity Gender-Session Interaction Plot

SDT Bias

The traditional SDT measures indicated that participants were liberal (i.e., tolerating fewer misses and more false alarms; Table 5). Of the fixed main or interaction effects analyzed, there were no significant effects, nor any trends.

Fuzzy SDT Sensitivity

As expected, the fuzzy SDT sensitivity measures were better than the measures from the traditional analysis (Table 6). Each of the 64 fuzzy SDT sensitivity measures were compared with the corresponding traditional measures using a two-sided paired Wilcoxon Signed Rank and the test was significant ($Z = -6.754$; $p < 0.001$). The better fuzzy SDT measures were due mainly to an improvement in the false alarm rate.

For the fuzzy SDT sensitivity measures, the gender effect was significant ($F_{1,7.658} = 7.608$ $p = 0.026$). As expected, the males (mean = 0.692; variance = 0.0019) performed better than the females (mean = 0.659; variance = 0.0019).

Fuzzy SDT Bias

As with SDT, the fuzzy SDT bias results indicated that all participants had a liberal bias (Table 6). The fuzzy measures were in general smaller (in absolute value) than the traditional measures, mainly due to false alarm rate differences. A two-sided paired Wilcoxon Signed Rank comparing the 64 fuzzy with the 64 traditional measures was significant ($Z = -6.504$; $p < 0.001$).

Table 6
Summary of Fuzzy SDT Results

	p(H)	p(FA)	d(A)	C
Mean	0.760	0.409	0.675	-0.245
Variance	0.0061	0.0071	0.0021	0.037

Of the main and two-way interaction effects, the only significant effect was for the session-noise interaction ($F_{1,16.105}$ = 3.557, p = 0.038) but there was a trend for the gender-noise interaction ($F_{1,11.159}$ = 3.530, p = 0.087). In the low noise condition, the bias remained relatively stable over time while in the high noise condition; participants became more liberal over time (Figure 5). With respect to the gender-noise interaction, the females were more liberal in low noise while males were more liberal in high noise (Figure 6).

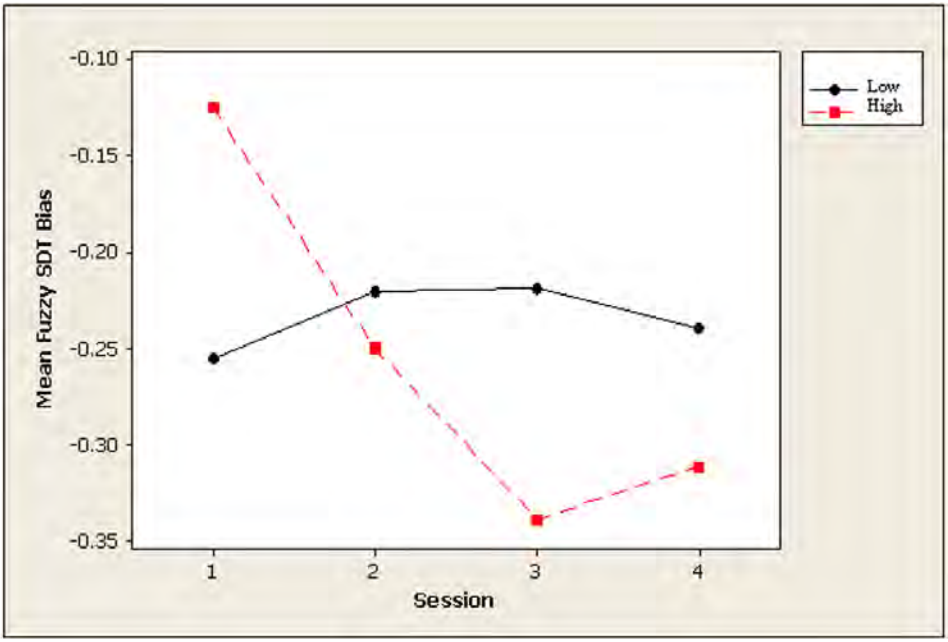


Figure 5. Fuzzy SDT Bias Noise-Session Interaction Plot

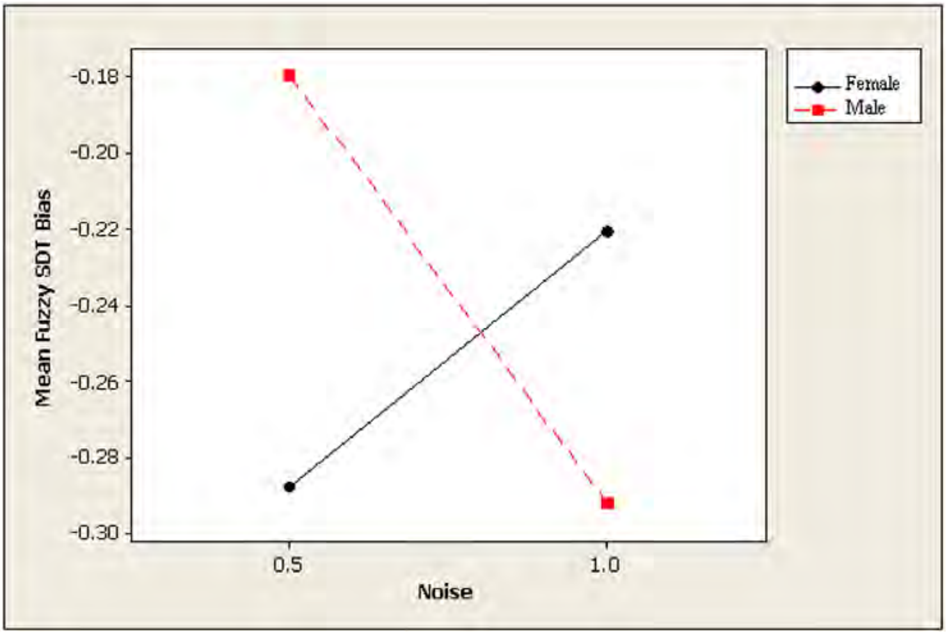


Figure 6. Fuzzy SDT Bias Gender-Noise Interaction Plot

Achievement

The achievement measures were low (mean = 0.270; variance = 0.039 based on the transformed probabilities). Some values were even negative. After applying the Fisher's r to zr transformation to the achievement values, the ANOVA indicated that only the gender effect was significant ($F_{1,12.514}$ = 5.584, p = 0.035). The achievement of the male participants based on the transformed probabilities (mean = 0.332; variance = 0.036) was superior to the females (mean = 0.208; variance = 0.036).

Linear Knowledge

Of the LME measures affected by human judgment, the G measures were the largest contributors to achievement (mean = 0.740; variance = 0.179 based on the transformed probabilities). This result was not surprising as G is generally high in judgment analysis research (Cooksey, 1996). None of the main and two-way interaction effects were significant nor were there any trends of interest.

Cognitive Control

The Rs values were low (mean = 0.343; variance = 0.028 based on the transformed probabilities). Of the main and two-way interaction effects analyzed using the transformed measures, the only finding was a trend for the gender ($F_{1,14.413}$ = 4.115; p = 0.061). The cognitive control of the male participants based on the transformed probabilities (mean = 0.383; variance = 0.033) was superior to the females (mean = 0.303; variance = 0.020).

Nonlinear Knowledge

The C measures were close to zero (mean = 0.013; variance = 0.033 based on the transformed probabilities). None of the main and two-way interaction effects analyzed using the transformed measures were significant or trends.

Skill score

For the SS, the data calculated by Equation 12 were analyzed using the transformed probabilities. On average, the scores were negative (mean = -0.253; variance = 0.183) meaning that the judgments were worse than the standard. With SS, the gender-noise interaction was significant ($F_{1,7.251}$ = 6.751; p = 0.034) (Figure 7).

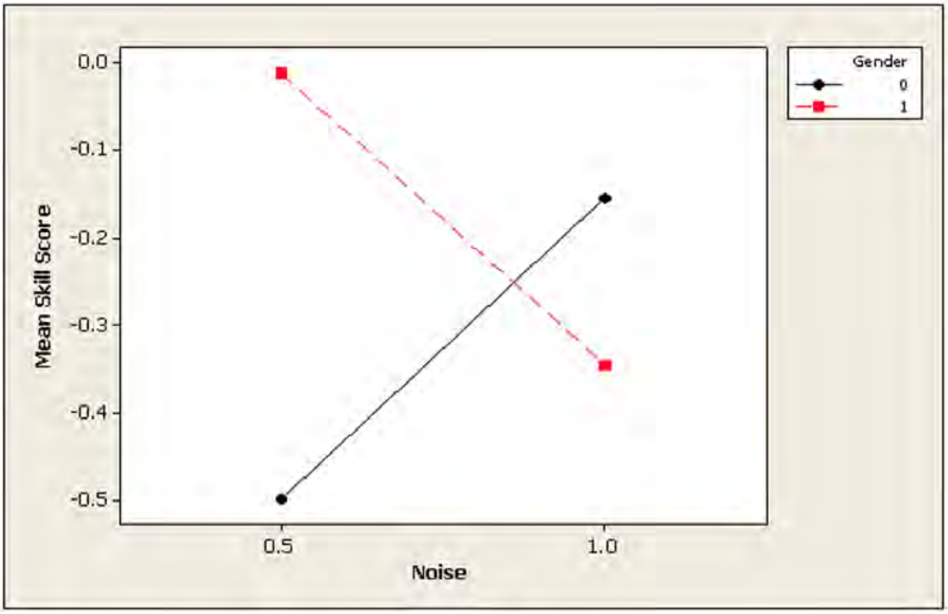


Figure 7. Skill Score Gender-Noise Interaction Plot

Conditional bias

The conditional bias measures indicated that the participants' judgments did not reflect the variability in the criterion (mean = 0.258; variance = 0.095). The participants' judgments tended to not cover the range that the criterion P(C) did. Of the main and two-way interaction effects analyzed using the transformed measures, there was a trend for the gender-noise interaction ($F_{1,10.845}$ = 4.236; p = 0.064) (Figure 8). The low noise males had lower (better) conditional bias than the high noise ones, while the high noise females' measures were better than the low noise ones.

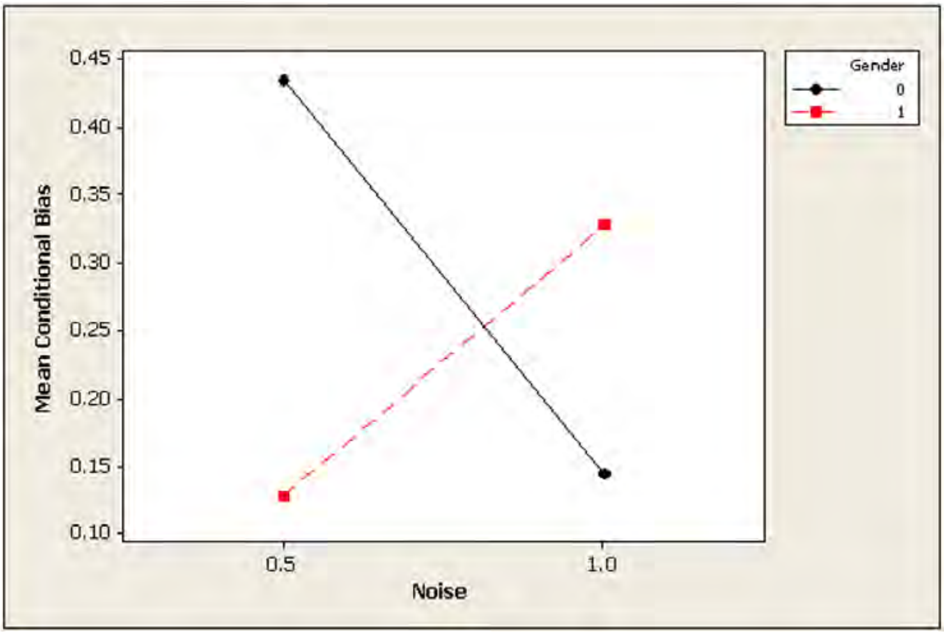


Figure 8. Conditional Bias Gender-Noise Interaction Plot

Unconditional bias

On average, the unconditional bias measures were positive (mean = 0.1062; variance = 0.017) thus diagnosing a tendency to over-estimate the P(C). Of the main and two-way interaction effects analyzed using the transformed measures, none were significant nor were there any trends of interest.

Discussion

The focus of this effort was to compare four quantitative methods of measuring human judgment performance for the same noisy experimental task. The focus was not on the results of a particular measure for the artificial task used in this study but rather on the similarities and differences between the measures. As the analyses were limited in the sense that only one cutoff value was used for the traditional SDT analysis (50%), only one of many possible environmental criteria was used for the other approaches, and the sample size in this experiment was small, the ANOVA analyses were not expected to find many statistically significant differences. However, interesting differences between the methodologies were identified and the analyses did highlight the notion that a single approach alone may not be sufficient for analyzing human judgment performance.

With respect to the four aggregate performance measures, traditional and fuzzy SDT, achievement and the skill score, the results were different. Both a fuzzy signal detection theory sensitivity measure and the lens model achievement indicated a significant gender effect, with males outperforming females.

However, neither SDT nor the skill score indicated this gender effect. Only the skill score indicated a significant gender-noise level interaction and only SDT indicated a trend for the gender-session interaction.

A comparison of the traditional with fuzzy SDT sensitivity and bias measures yielded different results. As expected, the fuzzy SDT sensitivity measures were better and the fuzzy SDT bias measures were smaller in absolute value. As mentioned, while the fuzzy measure indicated a gender effect, the traditional SDT sensitivity measure indicated a trend for the gender-session interaction. While both methods indicated a liberal bias in the participants, only the fuzzy bias measures indicated a trend for the gender-noise interaction and a significant noise-session interaction.

At a deeper level of decomposition, the MSE-based and the regression-based approaches also indicated alternative between-subjects differences with respect to judgment performance. The cognitive control measure, Rs, indicated a trend for the gender effect in the same direction as achievement and the fuzzy SDT sensitivity measure. Similar to the fuzzy SDT bias measure, the conditional bias measure indicated a trend for the gender-noise interaction where in both cases, males had better values in the low noise condition, and females had better values in the high noise condition.

While the experiment was based on an artificial task using an artificial criterion and while there was a small sample that may have allowed individual difference to influence the results, the comparisons still highlight the need for analysts not to rely on a single measure in analyzing human judgment performance in real-time judgment tasks. The signal detection, MSE-based and the regression-based approaches alone cannot individually shed light on the set of contributors human judgment performance.

The choice of methodology is left to the analyst. The analyst must consider what level of commitment can/should be made with respect to the criterion. For example, is it binary or is it continuous? If it is continuous, the analyst may gain better insight into judgment behavior by considering the methodologies that allows such constructs. Similarly, what level of commitment can be made with respect to the cues utilized by the judge and are the cue values available? If the analyst has access to the cue values, then judgment analysis techniques can be applied. If multiple methods are possible, the analyst should consider multiple methods as each can highlight different behaviors.

Future work should investigate what specifically can be learned about performance from each of the models. For example, under what conditions would findings be expected to be consistent or contradictory across the models? What types of independent variables might be expected to cause differences in model parameters? In this way, analysts would be better equipped to analyze human judgment performance.

References

Bass, E.J. (2002). Human-automated Judgment Learning: A research paradigm based on Interpersonal learning to investigate human interaction with automated judgments of hazards. *Dissertation Abstracts International*, 63 (03) B, (UMI No. 3046873).

Bolton, M. & Bass, E. J. (2005). Cognitive Systems Engineering Educational Software (CSEES): Educational software addressing quantitative models of performance. *2005 IEEE International Conference on Systems, Man, and Cybernetics.* Hawaii, October 10-12, 2005.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* (2nd ed.) Berkeley: Univ. of Calif. Press.

Campbell, G. E., Buff, W. L., & Bolton, A. E. (2000). The diagnostic utility of fuzzy system modeling for application in training systems. *In Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society,* 370-373.

Caplan, P. J. MacPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. *American Psychologist,* 40(7), 786-799.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and application*. New York: Academic Press.

Green, D. M. & Swets, J. A. (1989). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Hursch, C. J., Hammond, K. R. & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability. *Psychological Review,* 71(1), 42-60.

Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development,* 56, 1479-1498.

Maccoby, E. E. & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford: Stanford University Press.

Montgomery, D. C. (2000). *Design and analysis of experiments (5th ed.).* New York: John Wiley & Sons.

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review, 116,* 2417-2424.

Parasuraman, R., Masalonis, A. J., & Hancock, P. A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*(4), 636-659.

Pritchett, A. R. & Ippolito, C. (2000). Software architecture for a reconfigurable flight simulator. *AIAA Modeling and Simulation Technologies Conference.* Denver, CO.

Sanchez, E. (1979). Medical diagnosis and composite fuzzy relation equations. In M.M. Gupta, R.K. Ragade, & R.R. Yager (Eds.) *Advances in Fuzzy Set Theory and Applications* (pp. 437-444). New York: North-Holland.

Stewart, T. R. (1990). A decomposition of the correlation coefficient and its use in analyzing forecasting skill. *Weather and Forecasting, 5,* 661-666.

Tucker, L. R. (1964). A suggested alternative formulation in the developments of Hursch, Hammond & Hursch and by Hammond, Hursch & Todd. *Psychological Review, 71*, 528-530.

Wickens, C. D., Gordon, S. E. & Liu, Y. (1998). *An introduction to human factors engineering.* New York: Addison Wesley Longman.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8,* 338-353.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Improving Pilot Procedure Following Using Displays of Procedure Context

Steven J. Landry
School of Industrial Engineering
Purdue University
315 N. Grant Street
West Lafayette, IN 47907-2023
slandry@purdue.edu
765-494-6256

and

Julie A. Jacko
Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology
and
Emory University School of Medicine
Atlanta, GA 30332
julie.jacko@bme.gatech.edu
404-385-5041

Abstract

Despite the prevalence of procedures in safety critical systems, there has been little focus on developing automated support for procedure following. This document describes a method for leveraging an existing design methodology – procedure context – to provide guidance on the content of displays for this purpose. Procedure context categorizes information used when developing the procedure, but which is typically absent from displays and otherwise unavailable to the operator. In this paper, the method is applied to a pilot flying an instrument approach procedure. After identifying elements of procedure context which could assist the pilot and which are not currently presented, a display was developed which incorporated these elements. A simulator experiment was then run which demonstrated the utility of these elements to the pilot, reducing lateral error and improving situation awareness.

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.

Introduction

The control of most systems in safety critical environments is highly proceduralized, and procedures have been cited as one of the most significant causes of accidents in aviation (Graeber & Moodi, 1998; National Transportation Safety Board, 1994), the nuclear industry (Marsden, 1996; Trager, 1988), and maritime (Perrow, 1984). Despite this, there has been little human factors work on the design and utilization of the procedures themselves until recently, when a number of researchers focused on the design of aviation checklists following a spate of accidents related to flight crew checklists (Degani & Weiner, 1990). Almost all of the work focuses on the use or misuse of checklists, and often either explicitly or implicitly assumes that non-compliance to procedure is necessarily a source of error (Burian, Barshi, & Dismukes, 2005; DeBrito, 2002; Park, Jung, Kim, & Ha, 2002). One notable exception is a discussion by researchers at Boeing of how pilots' compliance, and the correctness of compliance or non-compliance, is dependent upon circumstances (Graeber & Moodi, 1998). Researchers have focused on the design of checklists (Degani & Weiner, 1993), on training to handle emergencies (Burian & Barshi, 2003), and the consistency of procedure guidance (Burian, Dismukes, & Barshi, 2003).

Procedures, despite their importance and ubiquity, are typically developed informally, pieced together from diverse requirements and constraints on the system (Degani, Heymann, & Shafto, 1999). The result is that the design philosophy behind a procedure, if there is one, is variable from instance to instance, complicating an operator's ability to execute the procedure. There is also almost no display or automation support provided to the operator for following or interpreting procedures. Most procedures are presented to the operator through training or written guides (such as checklists), which often lack information about the assumptions made when the procedure was designed. If the procedures themselves are under-defined or limited in scope, typically the checklists fail to reflect the circumstances under which the procedures apply. Some situations in which the operator could find him- or herself have been considered by the procedure designer, but some have not. Without this knowledge, the operator is in a poor position to determine whether the procedure is still valid or not.

In this study, the authors identified elements of context of a procedure, which could assist pilots in following instrument approach procedures, and then tested whether these elements would result in improved situation awareness and safety when displayed dynamically. The identification of these elements followed from a description of procedure context given by Ockerman and Pritchett (Landry & Pritchett, 2002; Ockerman & Pritchett, 2000, 2004). This context information was used or assumed when designing the procedure, but was not provided explicitly to the pilot executing the procedure.

Background

Instrument approach procedures

Pilots flying under IFR and approaching an airport are required to fly an instrument approach procedure (IAP), the rules for which are identified in the Federal Aviation Regulations (FARs) (Federal Aviation Administration, 1994). These approach procedures are designed to transition a pilot from enroute airspace to the runway environment. Many approach procedures are designed so that even a minimally equipped aircraft could fly the approach. The details for a particular approach are contained in an instrument approach plate, a paper chart that identifies the navigation aid to be used, the runway to which the IAP is directed, the course to be captured and flown, and a descent profile. An example is shown in Figure 1.

In non-radar environments (or when told to by the air traffic controller), pilots must fly an IAP that takes them from any altitude and transitions them to the final approach. This procedure is designed to keep all aircraft flying approaches to the airport (and departures from the airport) separated, to keep the aircraft away from terrain obstructions, and safely and consistently turn the aircraft on to a stabilized final at an altitude that will allow a safe descent to the runway.

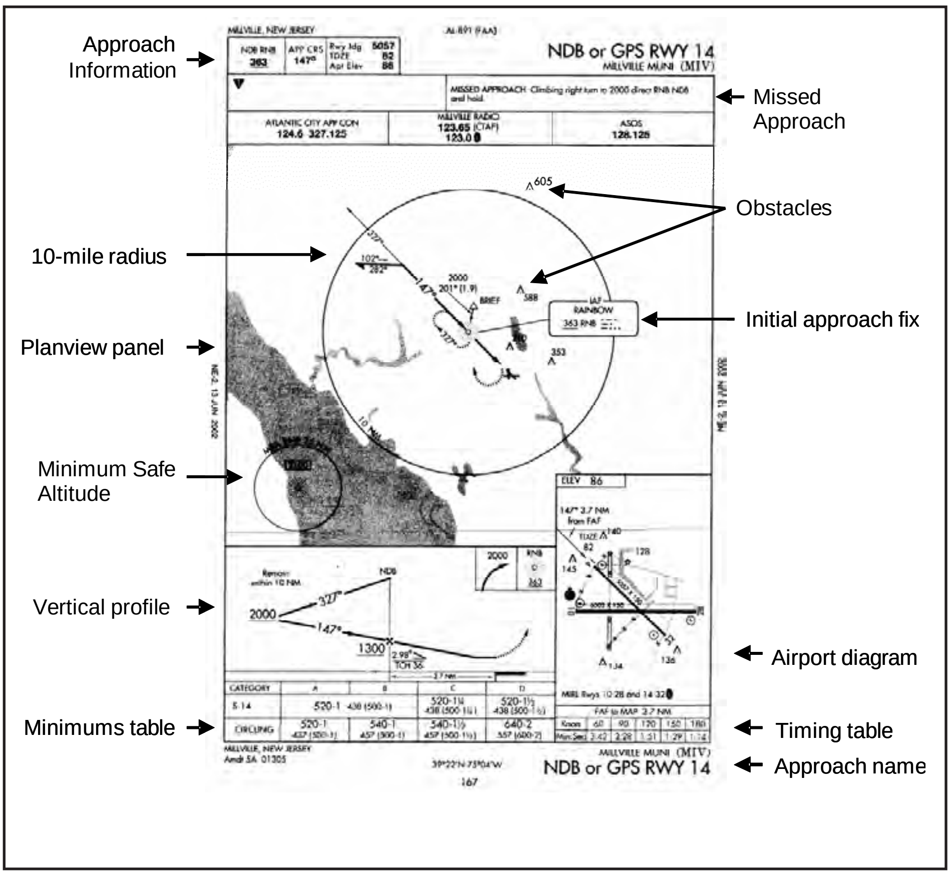


Figure 1. Instrument approach plate (Federal Aviation Administration, 2003).

To fly an IAP, pilots first cross the initial approach fix (IAF), then fly some track in order to lose altitude. Once crossing the final approach fix (FAF), pilots descend to the minimum descent altitude (MDA) (or decision height if glidepath guidance is available). Once reaching the MDA, pilots fly inbound until the runway environment is sighted or until the missed approach point (MAP), which is typically identified by either timing or by reaching a specified distance from a radio navigation aid. If the MAP is reached and a safe landing cannot be made, the pilot must "go around" – initiate a missed approach using the procedure specified on the approach plate. There are several types of IAPs, but there are many common elements on the approach plate. Referencing the callouts in Figure 1, the following elements are found on all approach plates.

- Approach Information (including the identifier and frequency of the initial approach fix -IAF, the approach course, and runway information) is located in the upper left corner.
- Missed Approach instructions are located in the upper right corner.
- A Minimums Table indicating the MDA and minimum weather requirements is located bottom left.
- A Timing Table, indicating how long it will take to get from the final approach fix (FAF) to the missed approach point (MAP) is located in the lower right corner.
- The name and runway of the approach are located below the timing table and above the missed approach instructions.
- An Airport Diagram, including the elevation, is located above the timing table;
- In the center of the approach plate is a planview display of the approach, (including a symbolic depiction of the approach, an indication of the IAF, indications of significant obstacles [which may be terrain or manmade], a circle around the IAF of a certain distance [in this case it is 10 nautical miles], and an indication of the minimum safe altitude [MSA – the altitude which provides 1,000 foot separation from the terrain within 25 nautical miles]).

Procedure Turn IAP

The procedure turn IAP has several segments. The approach plate has all the common elements of any instrument approach plate (discussed above). There are two additional elements: the "barb" and the "remain within" distance. The barb indicates the side of the inbound course on which the course reversal maneuver is to be performed. The "remain within" distance indicates the extent within which the course reversal must be accomplished.

These elements, in conjunction with the altitude restrictions provided on the approach plate, are chosen such that an aircraft complying with those restrictions will be more than 1,000 feet above any terrain feature until after the FAF.

The FARs state only that the procedure turn depiction must show "the outbound course, direction of turn, distance within which the turn must be completed, and minimum altitude" (Federal Aviation Administration, 1994). This means that, for a procedure turn depicted with a barb, the point at which the turn inbound is made, and the type and rate of the turn, is at the pilot's discretion. Typical maneuvers taught (and commonly used) are the 45º-180º maneuver, the racetrack pattern, the tear-drop, or the 80º-260º maneuver. These maneuvers are shown in Figures 2a through 2d. The headings indicated on the approach plate correspond to the 45º-180º maneuver, but this maneuver is not required.

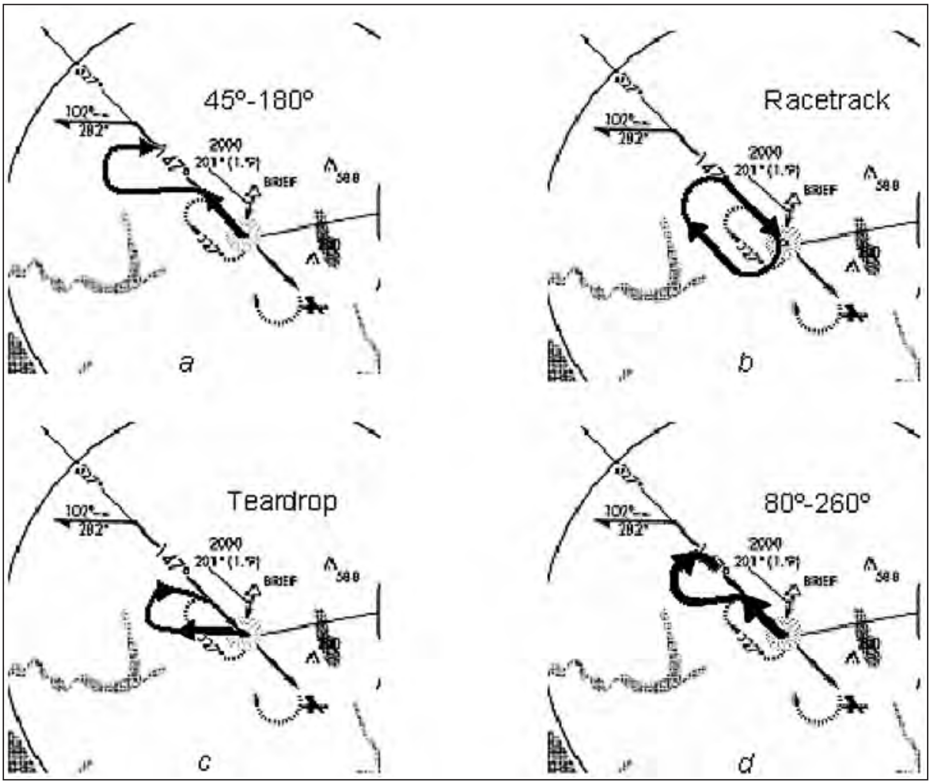


Figure 2. Entry maneuvers for procedure turn IAP.

The 45º-180º maneuver consists of a turn outbound (away from the airport) to parallel the desired inbound course for approximately 45 seconds to one minute, then a 45º turn towards the protected side (remaining on that heading for approximately 1 minute), then a 180º turn in the opposite direction as the last turn. This 180º heading should give the aircraft a 45º intercept heading to the inbound course. The aircraft then intercepts and follows the inbound course (and the remainder of the procedure).

The racetrack pattern is a technique typically used when the aircraft is already flying close to the inbound course to the procedure turn fix. In this case, it would take a 180º turn to start a 45º maneuver, and the aircraft would be well off the inbound course outbound by the time such a turn is completed. Instead, the aircraft simply turns 180º towards the protected side and flies that course outbound for one to one and a half minutes. The aircraft then turns 180º again (in the same direction), which should put the aircraft back on the inbound course.

The tear-drop maneuver consists of turning immediately to the 45º heading depicted on the approach plate once passing over the procedure turn fix, and flying that heading for one to one and a half minutes, then turning 135º back towards the inbound course. These turns should result in the aircraft being close to on-course inbound.

The 80º-260º maneuver is similar to the 45º maneuver except that the 80º heading is not maintained. Once the heading is obtained, the turn back towards the inbound course is commenced. This maneuver should result in the aircraft being close to on-course inbound.

The "remain within" distance must be complied with for terrain avoidance reasons. Outside this region, 1,000-foot obstacle clearance is not guaranteed if the aircraft is below the MDA. The standard maneuvers described above have been shown to remain within 10 miles. If the pilot uses any other maneuvers or if the "remain within" distance is 5 miles, then some planning is required, and typically different times for when to turn need to be used in order to stay within the protected airspace.

Procedure Context

Ockerman introduced the concept of procedure context to categorize information used to design a procedure. Much of this information is not passed to the operator who implements the procedure, although this information could be useful for understanding the underlying assumptions and constraints of a particular procedure. Ockerman's taxonomy contained two main categories of procedure context information – explanatory and locational, and their elements.

Explanatory context provides background information on the procedure, indicating purposes and interrelationships within the procedure. It helps the user apply a strategy to accomplish the procedure, and aid him or her in understanding consequences of not complying with the procedure, or in enabling him or her to safely alter the procedure during execution. Normally much of this information is not transferred by the procedure designers and is lost. If retained, however, the information can be provided to the user through training, documentation, or displays. Elements of explanatory context are:

- Intention –the overall goal state of the procedure.
- Rationale –the reasons for individual steps.
- Boundary conditions –the conditions under which the procedure is assumed to be operating.
- Triggering conditions –external conditions that cause a procedure to begin, branch, or end.
- Temporal construct –the time window in which the procedure is assumed (or is required) to be accomplished.
- Ordinality – the requirements for the order in which steps must be accomplished.
- Necessity –the degree of requirement that the step be accomplished.
- Reversibility –the degree to which actions accomplished as part of the procedure can be "taken back."
- Appropriate specificity –the degree to which the procedure captures the detail of what needs to be accomplished.

Locational procedure context is intended to provide information concerning the physical ordering of the procedure. Its elements are:

- Previous actions – this relates to the actions that have been already accomplished.

- Following actions – this relates to the actions that are upcoming.
- Location indication – this relates to where in the global procedure the current step resides.
- Forking – this relates to how a procedure might branch.

Display Design Guidance: Identifying and Using Procedure Context

Identifying procedure context elements for procedure turn

The regulatory guidance provided for procedure turns was discussed above, and pilots only have reference to their training and the approach plate while executing the procedure. However, there is considerable context about the procedure to which they do not have access (and for which they probably have not been trained). In addition, the complexity of the procedure may make some of the context of the procedure, for which a pilot has been trained, inaccessible. Following is a description of each of the elements of procedure context as it relates to procedure turn IAPs.

Intention. The intention of the procedure turn IAP is to transition the aircraft safely and consistently from enroute airspace to a point at which a descent to landing can be made.

Rationale. Each segment of the approach can be considered a step of the IAP. For each segment, it is important that the pilot understand the rationale for that part of the IAP. Table 1 outlines the major segments of the procedure turn IAP, and the rationale for that step.

Boundary Conditions. Procedure turn IAPs are designed in accordance with FAA Order 8260.3B, titled “United States Standard for Terminal Procedures” (TERPs) (Federal Aviation Administration, 1976). This document mainly specifies the airspace restrictions within which the procedure must be confined, commonly referred to as the “protected airspace.” These restrictions form a significant portion of the boundary conditions for procedure turn IAPs, but do not form a significant part of training, pilot reference documents, or displays. Figure 3 indicates the protected airspace for a procedure turn.

Remaining clear of terrain obstacles is one of the main goals of IAPs, and this is guaranteed only when the pilot (1) remains within the airspace restrictions identified for the approach, and (2) remains above the specified altitudes for the particular portions of the approach. The latter is specifically identified on the approach plate and is (fairly) easily confirmed on the aircraft’s instruments. The former, however, is not clearly identified on the approach plate nor easily confirmed on the instruments.

Table 1
Rationale procedure context for procedure turn IAP.

Segment	Rationale
Cross the IAF	Marks beginning of approach and identifies a place to begin turn outbound.
Turn outbound	The fly-off outbound enables the pilot to lose altitude and get set up for a consistent intercept of the final approach course.
Inbound turn altitude restriction	The minimum altitude that keeps aircraft separated from terrain until on final.
FAF altitude restriction	Keep aircraft above terrain until established on final approach.
Timing	The MAP is identified as a distance from the FAF, but no distance measuring equipment is available; timing is therefore used to identify the MAP.
MDA	The minimum altitude that keeps aircraft on final separated from terrain.
MAP	This is a point at which a safe landing can no longer be made from the MDA.

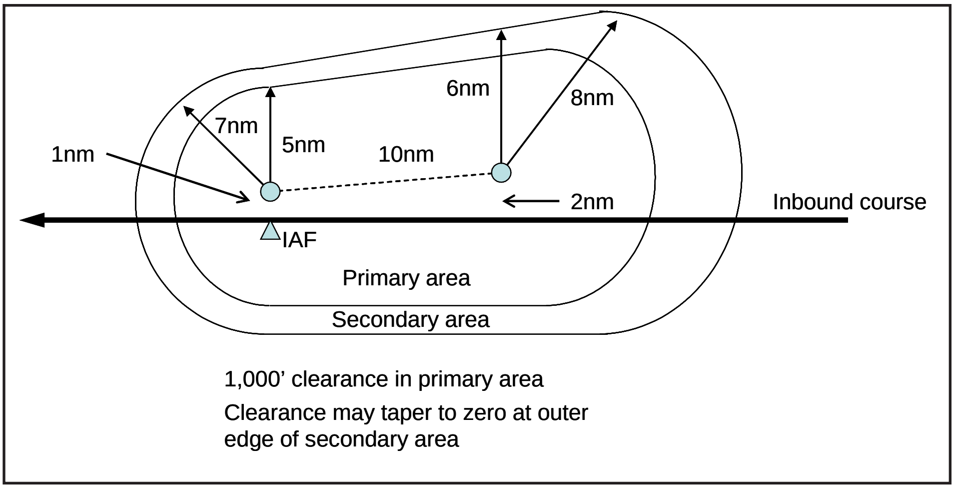


Figure 3. Protected airspace for IAP.

Triggering Conditions. Each segment in Table 1 has a triggering condition, as shown in Table 2. These conditions can be found in the regulations, but are not explicitly presented on the instrument approach chart or on any of the flight deck displays. The triggers indicate the events that must occur in order to transition from one step or segment to the next, or the criteria that, when met, releases the pilot from the restriction.

Table 2
Triggers for IAP segments.

Segment	Trigger
Cross the IAF	Station passage as indicated on navigation instruments
Turn outbound	After station passage
Inbound turn altitude restriction	Until established on the final approach course
FAF altitude restriction	Until passing FAF
Timing	Start at FAF
MDA	Maintain upon reaching until runway in sight or MAP
MAP	Upon expiration of timing

Temporal Construct. Due to the minimum navigation instruments required for the procedure turn IAP, the pilot does not necessarily have access to any direct information regarding distance from the IAF or FAF. Since the aircraft is (in this case) instructed to remain within 10 miles of the IAF, and needs to know when the MAP is reached (which is only identified by it being 3.7 NM from the FAF), timing must be used in lieu of distance measurements. Pilots remain within 10 miles by flying away from the IAF for only a predetermined amount of time (generally 1 to 2 minutes). Pilots identify the MAP by using the timing block on the approach plate, which identifies different times for a number of groundspeeds.

The groundspeed for the approach differs depending on the aircraft type and pilot technique, and also will vary over the course of the approach due to imprecision and winds. As a result, the actual distance flown for a given time will vary.

For the fly-off from the IAF, it is only important to remain within 10 NM, but it is nearly impossible for the pilot to be sure of what distance has been flown without distance information. To determine this, the pilot would mentally integrate the distance flown given the variation in ground speed. Of course, an estimate could be made based on an average speed and time. For an aircraft flying 120 knots groundspeed (2 miles per minute), 2 minutes of flight would result in the aircraft traveling 4 miles. If this estimate is off by 20 knots, the distance estimate would be off by 2/3 of a mile. This is not significant unless the limit is to stay within 5 NM (which is usually common for approaches used by general aviation aircraft). For faster aircraft, 2 minutes at 4 miles per minute (240 knots) is 8 miles, and if off by 20% the aircraft would be within 2/5 NM of exceeding the 10 NM limit.

For identifying the MAP, pilots start timing when passing the FAF, and would typically use the time associated with the expected final approach speed. If the expected final approach speed is not one of the entries on the timing table, pilots typically use the closest time or interpolate (often crudely) to get a closer time. This estimate becomes less accurate the more the actual groundspeed differs from the expected final approach speed.

Without any automation support, however, it is impractical for pilots continually to adjust fly-off times based on actual groundspeed, to estimate groundspeed, or to interpolate accurately fly-off times.

Ordinality. Certain steps are required prior to initiation of others. For the procedure turn IAP, the sequence is generally obvious – it is either illogical or physically impossible to do some steps before others. For example, it is illogical to perform a course reversal after passing the FAF. However, it may also seem illogical to turn back outbound if the final approach course is intercepted, yet this is required once the IAF is passed. The reason for the outbound turn in this case is not to better align the aircraft with the final course but to enable the pilot to lose altitude and also to keep separation with other aircraft on the approach ahead of the pilot. The ordinality constraint here should be identified, and, if possible, the reason for the constraint should be made clear.

An example of where ordinality is occasionally violated is on the inbound turn and its associated altitude restriction. The aircraft must be established on the inbound course prior to descending below that altitude. Descending prior to establishing on course inbound may be dangerous in that full terrain separation assurance may not be provided. Yet it is common for pilots to descend sooner than is allowed by the procedure.

Necessity. For the procedure turn IAP, the only necessary items are the crossing of the IAF, the direction of turn, the altitude restrictions, the inbound course, the crossing of the FAF, and the identification of the MAP. Other aspects of the approach are techniques for complying with these necessary items.

Reversibility. Errors during the IAP may or may not be reversible. Dropping below an altitude restriction can be reversed by climbing back above it, assuming terrain does not intervene. Course errors are (hopefully) always being reversed. If position errors become extreme, however, it may not be possible to reverse them and resume the approach.

If navigation errors become large, re-intercepting course in time to complete the approach may be too difficult to complete safely. In this case, the pilot should abandon the approach. The missed approach procedure gives guidance on how to abandon the approach and typically relies on being on or near the approach course.

Appropriate Specificity. Much of the procedure turn IAP is non-specific; numerous methods (or "techniques") are available to accomplish the procedure. This ambiguity is deliberate – the entry may be different depending on the approach to the IAF, and it is desirable to accommodate a range of techniques whose relative merits can be evaluated by the pilot under her or his specific circumstances.

The altitude restrictions depict their specificity. In the example provided, the restrictions are "at or above," which is indicated by the line below the altitude. If the restriction were meant to be "at" only, then there would also be a line above the altitude. Similarly, if the restriction were meant to be "at or below," then there would only be a line above the altitude. Altitudes that are simply recommended appear with no lines (Federal Aviation Administration, 2002).

The timing block appears specific when in fact it is not. Fluctuations in groundspeed and variations in where timing is started (relative to the FAF) will change the time it takes to fly to the MAP. However, without guidance that is more specific, the pilot must rely on the timing estimate.

Previous Step and Following Actions. The sequence of the IAP is fairly well defined, but many individual items need to be accomplished at nearly the same time. For instance, when departing the FAF, pilots need to begin timing, they may need to turn to a new final course, they need to establish a descent rate, and they may need to make a radio call to announce passing the FAF. Omitting any of these steps can lead to dangerous or unsafe situations.

Locational Indication. The procedure has two interconnected axes – the lateral profile of the procedure and the vertical profile. Pilots, with difficulty, can identify their position on the procedure for these two axes. The workload to do so, however, is typically higher than pilots can manage while also flying an instrument approach. Pilots flying the approach shown in Figure 1 (assuming they do not have access to a GPS display) have only an NDB to determine their position. They would be significantly taxed to identify their position with respect to the procedure airspace while manually controlling the aircraft.

The NDB approach shown in Figure 1 is the IAP with the least informative navigation instruments. Most approaches utilize better navigation aids and instruments. In these cases, pilots would have better locational procedure context, which may contribute to those approaches being less error-prone.

Forking. Forking in the procedure occurs when the aircraft must go around. At that point, a transition to the missed approach procedure must be accomplished. Pilots must go around if, when reaching the MAP, the runway cannot be seen or a safe landing cannot be made, anytime it is determined a safe landing cannot be made, or when instructed to do so by air traffic controllers (Federal Aviation Administration, 2002).

As mentioned previously, the MAP on the NDB approach is identified by timing, which makes its identification imprecise. Yet the missed approach procedure is designed with the assumption that the aircraft is at the MAP. Beginning the missed approach procedure away from the MAP can affect the ability of the aircraft to complete the procedure safely or cause separation problems with other aircraft.

Incorporating procedure context into display elements

Having identified the procedure context for the procedure turn IAP, one can then identify useful content for displays to support procedure following. Table 3 shows the context elements and how those elements are currently made available to the pilot.

Table 3
Procedure context elements in the procedure turn IAP.

Element	Where found in procedure turn IAP
Intention	Training
Rationale	Training, FARs, TERPs, Approach plate
Boundary conditions	TERPs
Triggering conditions	Approach plate, instruments, visual contact with runway
Temporal construct	Clock, timing block, groundspeed
Ordinality	FARs, Training, Approach plate
Necessity	FARs, Approach plate
Reversibility	Instruments
Specificity	Approach plate, FAR
Previous/Following actions	Training, FAR, Approach plate
Location indication	Instruments, Approach plate
Forking	Instruments, Approach plate

Of the elements in Table 3, several could benefit from being represented in some way to the pilot other than the manner in which it is currently done. The rationale for remaining within 10NM, for the turn to be conducted on a particular side, and for altitude restrictions is given by TERPs criteria. However, only minimal elements of this rationale are provided to the pilot on the approach plate. A more specific depiction of this rationale, which could be accomplished by explicitly identifying the protected airspace, would be extremely useful for the pilot to see when selecting maneuvers or when the procedure can no longer be followed in the standard manner.

The temporal constraints on identifying the MAP are given by the appropriate groundspeed entry on the timing block on the approach plate, and referenced by a clock. This is not ideal because the actual position of the MAP is based on distance and not time. In order for the pilot to convert to time, a particular groundspeed must be chosen (although actual groundspeed is dynamic). The pilot must then start the clock at the correct location, and identify which time matches the entry for that groundspeed in the timing block. If the approach groundspeed falls between entries, interpolation may be required. For identifying the MAP, this series of estimates and somewhat mentally taxing operations could be replaced with a more specific, more informative, and more accurate countdown to the FAF. One simple improvement would be to perform the interpolation for the pilot, and display the time left to the MAP.

Altitude restrictions appear on the approach plate, but much of the context (necessity of the altitude restriction, triggering conditions for passing an altitude restriction) is not clearly provided to the pilot. Visual triggers of different vertical phases of flight and confirmation of the necessity of those restrictions could be identified by highlighting the currently appropriate altitude restriction.

Other than through a difficult mental transformation of information on the navigation instruments, the pilot's location within the procedure is not clearly specified anywhere on the flight deck. A location indication of the aircraft's position within the procedure should therefore be useful to the pilot.

Since much of the above information is dynamic, it cannot be solely addressed through training or a static presentation of the information (such as on an approach plate). The information, then, may either be presented on an existing dynamic display, or presented on a newly developed display. For experimental purposes, it is desirable to be able to isolate the effects of adding the context information; minimizing the use of new displays and new symbology is therefore preferred.

There are three distinct items described above: timing, altitude restrictions, and procedure location. The first two may be simply displayed as text or graphically on existing instruments. The last requires a display similar to the approach plate itself, which would require the addition of a new map-like display.

What must be determined, then, is how to display timing information, altitude information, horizontal location information, and procedure location. For an operational display, human factors and HCI principles would be applied to design these displays. For purposes of this study, the specific design considerations for an operational display were left to future studies. The distinction that is to be tested is whether the addition of displayed elements of procedure context is useful to the operator; as such, a display that effectively isolates these elements is needed. The resulting display may not be ideal from a human-computer interaction perspective, but needs to provide visibility as to whether the operator is benefiting from the addition of these display elements as opposed to anything else that is being presented.

Incorporating specific procedure context elements in a display

From this analysis, the elements for the experimental display were designed to address the following shortcomings of the paper-based approach plate:

1. No rationale for remaining within 10 nautical miles, for flying mainly on the protected side, or for meeting the altitude restrictions.
2. Poor representation of the temporal constraint required to identify the missed approach point (MAP) on the approach (the MAP could only be identified by timing).

3. No explicit representation of the triggers which allow the pilot to transition from one altitude restriction to the next.
4. No representation of location within the procedure.

Figure 4a shows the original paper-based approach plate. Figure 4b shows the baseline electronic version created for this experiment, which was intended to mimic the original paper-based approach plate. Figure 5 shows the electronic version enhanced with procedure context information designed to address the shortcomings listed above.

For the enhanced display, an indication of the protected airspace was added (for item 1 in the list above), a time-to-MAP countdown clock was added (for item 2), a highlight box for the appropriate altitude restriction was added (for item 3), and a highlight of the current segment of the approach was added (for item 4).

The enhanced display did not assume the availability of any information other than that available to the other instruments in the aircraft (e.g. it did not use absolute position information such as provided by GPS). This was to ensure that any benefit provided by the enhanced elements was not due to the addition of information about the environment.

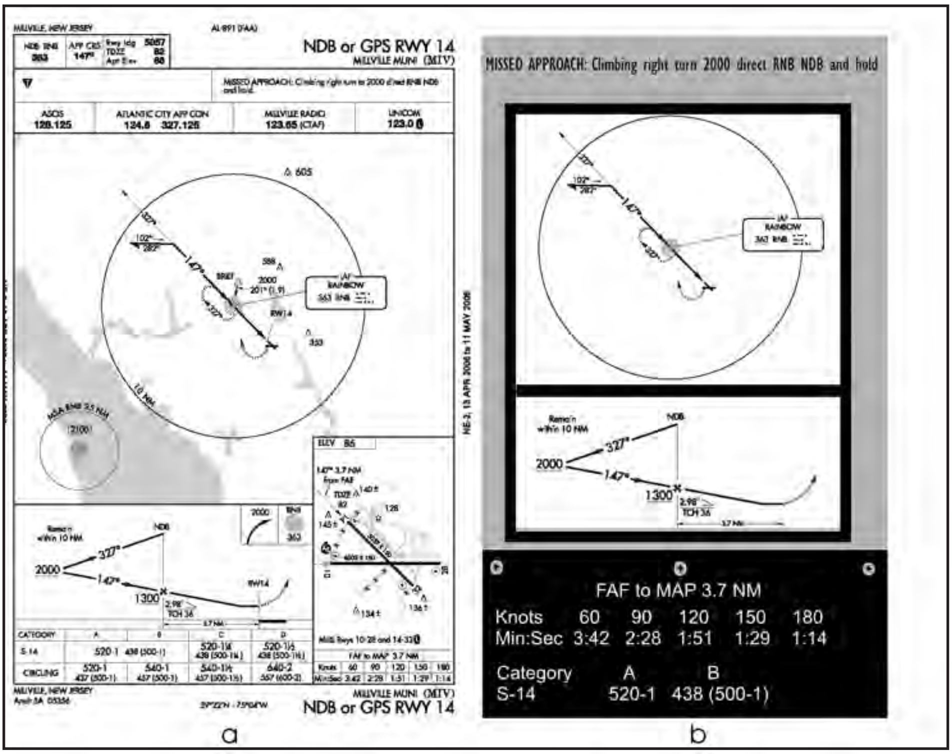


Figure 4. (a) Original paper based procedure display and (b) baseline electronic version.

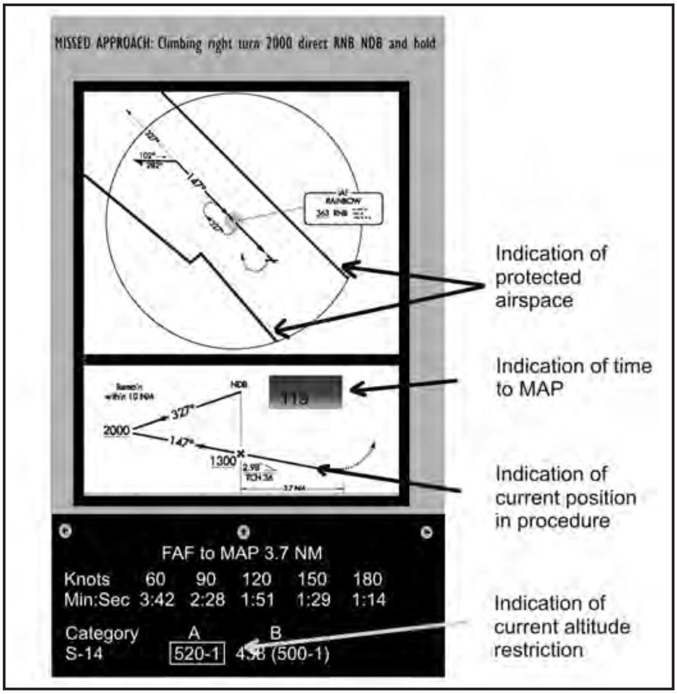


Figure 5. Enhanced display.

Experimental considerations

It is important to note that the purpose of the research was to test the belief that elements identified through an analysis of procedure context (such as was described above) could support procedure-following. It was not the purpose of the research to develop an operational display. As such, experimental considerations took precedence over an implementation that would be operationally viable.

For example, an operational implementation would be expected to reduce workload. For this experiment, however, a reduction in workload would confound with the display elements when attempting to analyze what might result in better performance and higher situation awareness. Therefore, the display was not designed to reduce workload. In fact, an ANOVA was unable to find an effect on workload due to display, although the power of the test was low ($\beta<0.8$).

Normally, the depiction of the IAP is accomplished using a paper instrument approach plate. Since the context of the procedure is dynamic, it is necessary to present context information in an electronic form.

Flight simulator experiment to test the addition of procedure context elements

Method

Using a desktop flight simulator, participant pilots flew a procedure turn IAP to Millville Municipal (an airport in New Jersey) using the baseline electronic approach plate in half of the trials, and the enhanced display in the other half. Their situation awareness, workload, and performance were then compared across the two display types. It was expected that the enhanced display would increase situation awareness and performance while not affecting workload.

To introduce complexity, the entry angle for the procedure turn was varied. The entry angles used were designed to provide the participant with clear guidance as to the initial turn required, ambiguous guidance, or conflicting guidance. An illustration of the entry angles is shown in Figure 6. One position (Figure 6a) was such that the aircraft approached the procedure turn fix conveniently aligned to perform a 45º-180º or an 80º-260º entry, so that a left turn to the outbound course began the procedure. From a second position (Figure 6b) pilots were conveniently aligned to perform either a teardrop entry or a holding entry, and could have either turned to parallel the course outbound, or turned 45º to the right to enter the protected airspace. From a third position (Figure 6c) no convenient entry maneuver existed, so that the pilot was unable to begin the procedure turn on the protected side unless some deviation from either ATC instructions or the instrument approach procedure was performed.

The experiment tested all combinations of the 3 X 2 test matrix. Each pilot flew two or three practice runs (until he or she felt proficient), then six data collection runs. Half of the pilots flew the runs with the baseline display first, and the other half flew with the enhanced display first. The order of the runs for the entry variable was random.

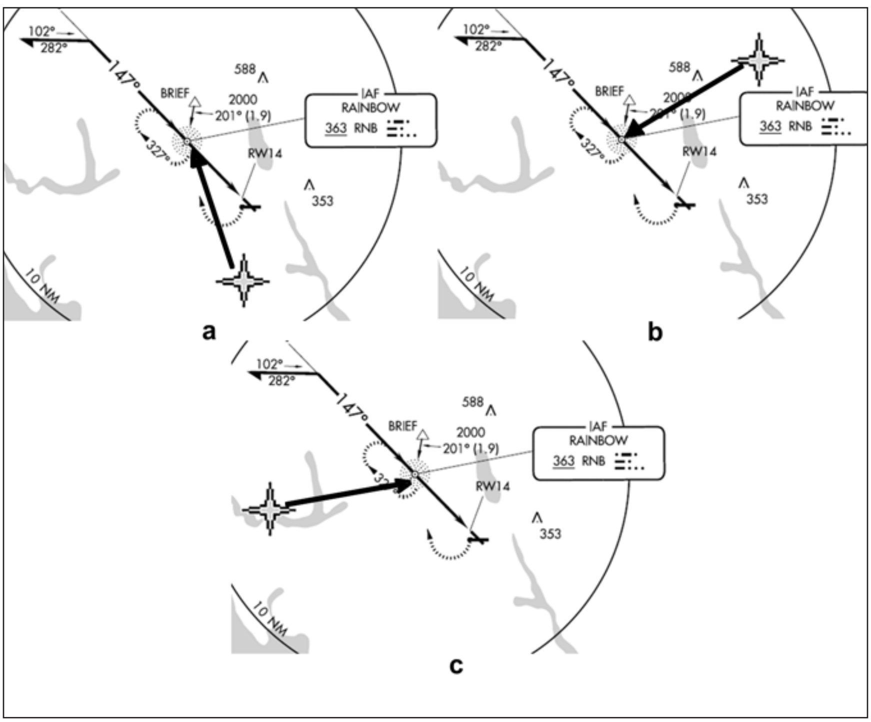


Figure 6. Initial positions and entry angles.

Each participant was briefed on the equipment and task, filled out a demographics questionnaire, then flew the practice and data collection runs. During each approach, aircraft state data was recorded. After each approach, a recall questionnaire regarding the entry and missed approach, and a workload questionnaire, were administered.

Equipment

The participants flew a desktop flight simulator running on a 1.2 GHz IBM Thinkpad R31 laptop computer. The software used was Microsoft Flight Simulator 2002, with the standard flight instruments for a Cessna 182. The baseline or enhanced display was added in place of some of the engine instruments (other engine instruments were still available on the opposite side of the panel). The instrument panel with the baseline display is shown in Figure 7.

A CH Products Flight Sim control yoke and throttle quadrant, and CH Products rudder pedals, were used to control the aircraft. The control yoke included controls for flap position, pitch and aileron trim, mixture control, and prop speed.

Figure 7. Simulator instruments with baseline IAP display.

The simulator was set for IFR flight to Millville Municipal Airport in New Jersey. To minimize additional flying difficulty (beyond that incurred by flying an unfamiliar simulator without motion), winds were set to calm. To ensure that the pilot relied solely on instruments to perform the procedure turn and approach, the weather was set for an overcast cloud layer at 400 feet above ground level, and with 10 miles visibility. The aircraft positions were preset as required by the experiment matrix, at an altitude of about 2,600 feet, and at a speed of about 120 knots.

Participants

Participants were 15 private pilots from the West Valley Flying Club in Palo Alto, CA. Subjects were paid $50 for their participation, and no incentives for performance were offered. The subjects were instrument trained and indicated they were familiar with the procedure turn IAP. Except for one subject (who was an advanced student recommended by his instructor), all subjects were instrument rated. As seen in Table 4, the subjects represented a wide range of flight hours, from a minimum of 152 hours to a maximum of 13,700. Three subjects were female and five were instructors.

Table 4
Subject demographics.

	Age	Total Hours	Instrument Hours
Median	**35**	**700**	**90**
Minimum	**26**	**152**	**30**
Maximum	**51**	**13700**	**3000**

Measures

Aircraft state data were recorded by Microsoft Flight Simulator then imported into Microsoft Excel, Minitab, and S-Plus. These packages were used to analyze the data. The state data recorded were aircraft position, speed, heading, altitude, and control actions. The state data was used to produce several composite measures, including starting position, position crossing the IAF, outbound track, intercept angle with desired track, inbound track, position crossing the IAF, final approach track, and position when starting the missed approach.

After each trial, a questionnaire designed to test the pilots' situation awareness of what transpired was administered. The subjects were asked to indicate on a map their initial position and ground track for the flight they just completed. Subjects' estimates of position as drawn on the map were compared with the aircraft's actual ground track for the composite measures listed above. The subject's recollection was scored as to being correct (1) or incorrect (0), and the sum of these eight represented the subject's score for recall of lateral position.

Next, the pilots were asked to indicate their recollection of any deviations from the altitude requirements, the speed requirements, or the missed approach requirements of the procedure, and the reasons for those deviations. These recollections were scored in comparison to the actual state data.

After the situation awareness questionnaires were completed, pilots were asked to fill out a NASA TLX workload questionnaire (Hart & Staveland, 1988).

The actual track of the flight was compared with the minimums for the approach to obtain a measure of flight error. Laterally the procedure required the pilot to remain on a bearing of 327º from the NDB when inbound to the FAF, then follow the 147º bearing out of the NDB until the MAP. Vertically the pilots were required to stay above 2000 feet until on the 327º bearing from the NDB inbound, stay above 1300 feet until the FAF, then stay above 520 feet for the remainder of the approach. The sampled position and altitude were compared with these limits for the appropriate parts of the approach to obtain a measure of error. Both the magnitude and duration of the error was calculated. In addition, when flight error was detected, the pilot's correction back to the procedure was calculated. This was intended to provide an indication of when the error was recognized by the pilot.

The number of control inputs was extracted from the state data. The values of the control input variables were plotted and summary statistics determined. More frequent and greater extent of control inputs are an indicator of increased workload, and provide a measure of the stability of the approach. Greater stability on the approach is desirable, and low stability is considered unsafe.

Results

Recall that the elements were added to address four specific shortcomings: (1) remaining within the safe confines of the approach, (2) properly identifying the missed approach point, (3) properly identifying when to transition from one segment of the approach to the next, and (4) properly identifying the current location within the procedure. The results described in this section relate to the specific elements added to the display.

Effect of display variable on pilot's ability to remain within the confines of the approach

For the extent of lateral deviations from procedure, a normal scores transformation normalized the data, which allowed an ANOVA to be run on the data. The results are shown in Table 5. In addition to the entry variable, the location on the approach, and an interaction between the entry and position on the approach, the enhanced display was also found to have a significant effect on the extent of lateral error. An examination of the data means (Figure 8) indicates that there was less error when using the enhanced display as compared to the baseline display, and an increased amount of error when executing the procedure from the conflicting entry position, particularly for the early portions of the approach (intercept and outbound).

Table 5

Analysis of Variance for Error_NormScore, using Adjusted SS for Tests.

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Subject	14	42.158	42.776	3.056	4.470	0.000
Display	1	3.473	3.306	3.306	4.830	0.029
Order	5	10.174	40.515	2.103	3.070	0.010

EntryNumber	2	4.520	2.865	1.433	2.090	0.125
Location	2	34.232	32.740	16.370	23.920	0.000
EntryNumber* Location	4	18.449	18.449	4.612	6.740	0.000
Error	295	201.867	201.867	0.684		
Total	323	314.872				

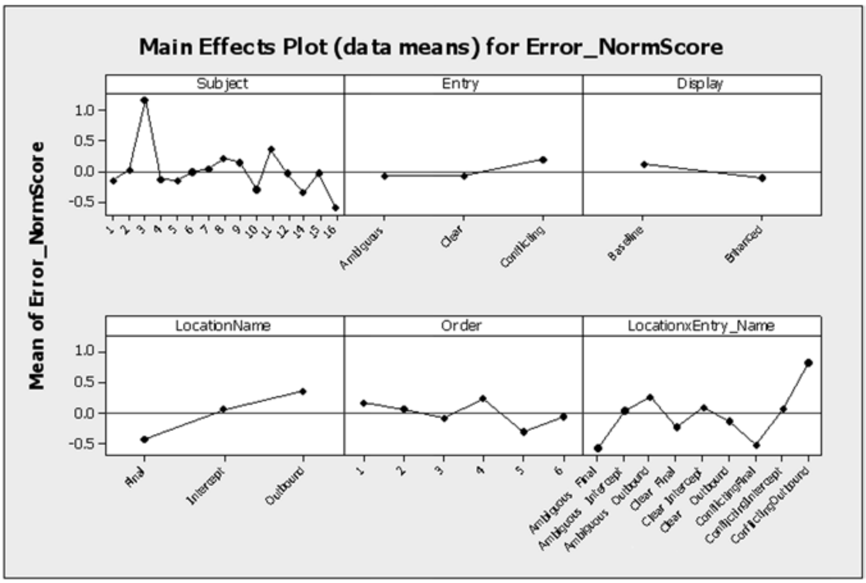


Figure 8. Differences in means for lateral error extent normal scores.

With one exception, all errors over 1.5NM on the non-maneuvering side of the approach airspace occurred with the baseline display. The exception was a case in which the subject was intentionally deviating from procedure; the subject had the same error while using both the baseline and enhanced displays. There were nine such errors involving five subjects, and seven occurred with the conflicting entry case. The largest error took the participant's aircraft 3.6 NM off course, 0.4 NM from the boundary of the primary obstacle clearance area.

Effect of display on the identification of missed approach point

An examination of the subjects' apparent identification of the missed approach point was compared to the actual missed approach point. Subjects executed the missed approach on average 0.5 nautical miles late (i.e. beyond the actual missed approach location). This was true regardless of display. Levene's test of equal variance, however, found that the variance of the missed approach execution point was significantly less (0.053 vs. 0.096) for the enhanced display ($F=3.545$, $p=0.063$).

After each simulation run, subjects were asked to recall missed approach deviations. The results were checked using the Kruskal-Wallis nonparametric test. No significant effect for subject, order, or entry was found to exist for the ability of the subject to recall missed approach deviations. However, a marginally significant effect (p=0.069) was indicated for display, as shown in Table 6 (higher scores indicated better recall).

Table 6
Kruskal-Wallis test on missed approach recall.

Display	N	Median	Ave	Rank	Z
Baseline	45	1.000		40.5	-1.82
Enhanced	45	1.000		50.5	1.82
Overall	90			45.5	

H = 3.30 DF = 1 P = 0.069

Effect of display on segment transitions

Since segment transitions are marked by changes in altitude, reductions in errors in vertical position would be indicative of improved segment transition compliance. A third root transformation normalized the extent of vertical error data and an ANOVA was run, but found no significant effects. Duration of vertical error data could not be normalized, but was tested using a Kruskal-Wallis nonparametric test; a marginally significant effect (p=0.077) for the display variable at the intercept location was found. A check of the sample means (Figure 9) indicated that, for these subjects, shorter durations for vertical errors occurred at the intercept location when using the enhanced display.

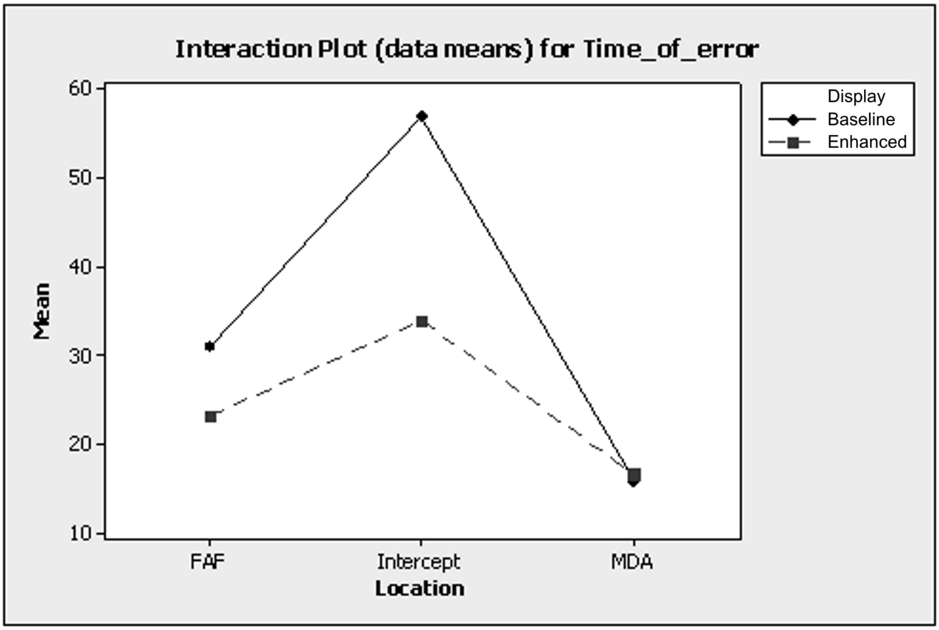


Figure 9. Mean duration of vertical errors by display and location.

Effect of display on identifying current location in procedure

Two measures of situation awareness were used to evaluate the subjects' ability to keep track of their position within the procedure. First, a questionnaire was administered immediately following each experimental run. This questionnaire tested subjects' recall concerning deviations from the constraints of the procedure for various locations. Secondly, the recall results were compared to an analysis of the duration of errors, with longer errors indicative of lower situation awareness.

ANOVA tests on the results of the recall tests indicate that subjects' recall for lateral deviations and vertical deviations was significantly improved when using the enhanced display ($F=5.69$, $p=0.02$ and $F=14.77$, $p<0.001$ respectively). In addition, as mentioned above, Kruskal-Wallis nonparametric tests on duration of vertical errors found a marginally significantly reduction when using the enhanced display.

Discussion of Results

The results indicate that the addition of procedure context elements assisted the pilots in remaining within the confines of the approach and eliminated nearly all large lateral errors. Significant improvements were found for the subjects' situation awareness, suggesting that his or her ability to keep track of location in the procedure was improved. Marginal results were found for improvements in situation awareness while executing the missed approach procedure and for decreasing vertical errors, suggesting that improvements may have been provided while suggesting that further experimentation should be done to confirm the results.

The reduction in lateral error found, when utilizing the enhanced display, is consistent with the elements of procedure context improving the pilots' ability to remain within the confines of the approach. Several aspects of procedure context information may have contributed to this improvement, and no attempt was made to isolate the effects of one (or combinations) of the elements. In addition, improvements in situation awareness also found as part of this experiment may have contributed to the overall reduction in lateral error.

The reduction in large lateral errors was dramatic, suggesting that the pilots' ability to remain situationally aware regarding their position on the approach was improved. This is further confirmed by the situation awareness results.

That a significant lateral effect but not an altitude effect would be found, is somewhat surprising since the enhanced display seems to provide more direct support for altitude compliance rather than track compliance. No concrete evidence indicates why the display did not support altitude compliance, but three possible reasons are presented here. First, it is important to note that the scale of altitude deviations (as compared to lateral deviations) is different. Altitude deviations are measured in 100s of feet, whereas lateral deviations are typically 1000s of feet or even miles.

Secondly, as mentioned earlier, the required altitudes are specifically identified on the approach plate, and there is an instrument that directly indicates altitude (the altimeter). Therefore, errors can be directly identified by the pilot with reference to one instrument and his recall of (or reference to) the approach plate, whereas lateral errors must be gleaned from crosschecking two or more instruments and mentally computing angular differences. The need for such "distributed" representations, where part of the problem is represented externally (on the instruments) and part is represented internally (angular computations), has been described by Norman (Norman, 1993), and can be extremely taxing for pilots, particularly during the period of high workload customarily associated with final approach. Overall, this makes altitude errors much more obvious than lateral errors, with or without procedure context elements.

In addition, the simulator's characteristics were different for the two axes. The simulated aircraft was very stable laterally, but not very stable in the pitch axis. This meant that corrections to lateral errors could be made and not tracked, whereas corrections to vertical errors needed to be continuously monitored. Since detection of deviations was not closely measured, it is not possible to prove this explanation. However, this type of control behavior could overwhelm any visible effect of the display on corrections to altitude deviations.

Conclusions

Procedures are required to operate safely many systems, yet little is provided in the way of support. The evidence from the experiment described above, in addition to the prevalence of procedures as a causative factor in accidents in safety-critical systems, suggests that such support should be provided.

This paper outlined in detail a method for identifying information which, when displayed dynamically, was demonstrated to provide benefits to pilots executing an instrument approach procedure. As such, it is intended to provide a guide for those wishing to design support for procedure following. Procedure context can provide a framework for describing information, which may be useful for following procedures, but whose presentation to operators is lacking or absent. These elements can then be added to displays to improve performance.

The displays described above are not ideal from an interface standpoint, but were designed to address experimental goals. As such, future work should address the specific form and content for interface elements such that overall performance and workload would be reduced. In addition, future experiments should attempt to identify reasons for the lack of statistical evidence showing an effect on vertical performance in this task. Finally, research should be undertaken to determine whether the results would carry over to training or non-dynamic aids.

References

Burian, B. K., & Barshi, I. (2003). *Emergency and abnormal situations: A review of ASRS reports.* Paper presented at the 12th International Symposium on Aviation Psychology, Dayton, OH.

Burian, B. K., Barshi, I., & Dismukes, K. (2005). *The Challenge of Aviation Emergency and Abnormal Situations* (No. NASA/TM-2005-213462). Moffett Field, CA: NASA.

Burian, B. K., Dismukes, K., & Barshi, I. (2003, August 2003). *The emergency and abnormal situations project.* Paper presented at the International Society of Air Safety Investigators 2003 Conference, Washington, D.C.

DeBrito, G. (2002). Towards a model for the study of written procedure following in dynamic environments. *Reliability Engineering and System Safety, 75*(2), 233-244.

Degani, A., Heymann, M., & Shafto, M. (1999). *Formal aspects of procedures: The problem of sequential correctness.* Paper presented at the 43rd Annual Meeting of the Human Factors and Ergonomics Society, Houston, TX.

Degani, A., & Weiner, E. (1990). *Human factors of flight-deck checklists: The normal checklist.* (No. 177549). Moffett Field, CA: NASA Ames Research Center.

Degani, A., & Weiner, E. (1993). Cockpit checklists: Concepts, design, and use. *Human Factors, 35*(2), 28-43.

Federal Aviation Administration. (1976). *FAA Order 8260.3B: U.S. Standard for Terminal Instrument Procedures (TERPS)*. Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration. (1994). *Federal Aviation Regulations Part 97: Standard Instrument Approach Procedures*. Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration. (2002). *Aeronautical Information Manual: Official Guide to Basic Flight Information and ATC Procedures*. Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration. (2003). U.S. Terminal Procedures (Vol. NE-2, pp. 167). Washington, D.C.: Federal Aviation Administration.

Graeber, R. C., & Moodi, M. M. (1998). *Understanding flight crew adherence to procedures: The Procedural Event Analysis Tool (PEAT).* Retrieved January 13, 2006

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.

Landry, S. J., & Pritchett, A. R. (2002). *Incorporating procedural information in a paired approach task.* Paper presented at the Human Factors and Ergonomics Society 46th Annual Meeting, Baltimore, MD.

Marsden, P. (1996). Procedures in the nuclear industry. In N. Stanton (Ed.), *Human Factors in Nuclear Safety* (pp. 99-116). London: Taylor and Francis.

National Transportation Safety Board. (1994). *A review of flightcrew-involved major accidents of U.S. air carriers, 1978-1990* (No. NTSB/SS-94/01). Washington, DC: NTSB.

Norman, D. A. (1993). *Things that make us smart*. Reading, MA: Addison-Wesley.

Ockerman, J. J., & Pritchett, A. R. (2000). A review and reappraisal of task guidance: Aiding workers in procedure following. *International Journal of Cognitive Ergonomics, 4*(3), 191-212.

Ockerman, J. J., & Pritchett, A. R. (2004). Improving performance on procedural tasks through presentation of locational procedure context: an empirical evaluation. *Behaviour & Information Technology, 23*(1), 11-20.

Park, J., Jung, W., Kim, J., & Ha, J. (2002, October 2002). *The step complexity measure - its meaning and applications.* Paper presented at the Korean Nuclear Society Autumn Meeting, Yongpyong, Korea.

Perrow, C. (1984). *Normal Accidents*. New York: Basic Books.

Trager, E. A. (1988). *Special study report: Significant events involving procedures* (No. AOED/S801). Washington, D.C.: Nuclear Regulatory Commission.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Retention of Aeronautical Knowledge

Stephen M. Casner
National Aeronautics and Space Administration
Ames Research Center
Moffett Field, CA 94035-1000

and

Daniel Heraldez and Karen M. Jones
San José State University Foundation
210 N. Fourth Street, Fourth Floor
San José, CA 95112

Abstract

Pilots' retention of aeronautical knowledge learned during private pilot training was studied in four experiments. In the first experiment, ten questions from the FAA private pilot airplane knowledge test were administered to sixty pilots, yielding an average score of 74.8%. Test scores were compared against seven characteristics of the pilots tested: certificates and ratings held, current role in aviation (pilot, CFI, or applicant for additional certificate/rating), total flight time, recent flight experience, reading habits, months passed since last evaluation, and months remaining until next evaluation. These factors explain some of the overall variability in test scores. Three follow-up experiments explored hypotheses about how retention might be affected by pilots' working environment: (1) pilots' knowledge becomes tuned to familiar aircraft charts; (2) difficult-to-remember regulations prompt pilots to substitute simpler rules that still allow them to operate legally; and (3) pilots' geographical region reinforces knowledge about local weather patterns, while knowledge of different weather patterns falls to disuse. The results well support two of these hypotheses but also further demonstrate that there are no simple-to-measure determinants of what aeronautical knowledge will be remembered and forgotten. The experience of everyday flying or teaching, together with recent flight experience and flight review requirements, does not appear to eliminate the need for ongoing study or rehearsal of aeronautical knowledge.

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.

Introduction

Learning to operate an aircraft requires the pilot to master a formidable amount of aeronautical knowledge. Knowledge about weather, regulations, aerodynamics, airspace, navigation, and aircraft systems and performance serves as the basis of pilot decision-making and actions. Mastering this aeronautical knowledge is known to be a laborious task, one that requires many hours of study (Flouris, 2001; Casner, Jones, Puentes, & Irani, 2003). In addition, after the pilot has initially learned this compendium of aeronautical knowledge, comes a second challenge: the challenge of remembering it.

It is well known that human memory is far from perfect, and it is natural that pilots will remember some of the things they have learned while they struggle to remember others. Hypotheses about what aeronautical knowledge is remembered and what is forgotten are easy to make. Memory research suggests that our ability to remember the things we have learned can best be summarized by a familiar adage: “Use it or lose it.” Specifically, the ability to retrieve any particular item from memory seems to be largely determined by how many times that knowledge has been used in the past, and how recently the knowledge has been used (Anderson, 1976). Therefore, it seems that everyday knowledge, such as aircraft performance and regulations that apply to routine flight, should be well retained; while less frequently used knowledge, such as emergency procedures and biannual inspection requirements, might fall to disuse. Of course, if our aim is to provide pilots with specific helpful advice, or perhaps to influence policy, we will need to be more systematic about making and validating our predictions.

One approach might be to attempt to catalogue which aeronautical facts and concepts pilots must recall and use throughout the course of their everyday activities, and to make predictions based on these usage profiles (Anderson, 1990; Newell & Rosenbloom, 1981). Unfortunately, for a domain as wide and rich as aviation, this process would seem to be both prone to error and laden with assumptions.

We adopt a more practical approach here by working the problem in reverse. We begin by capturing some of what pilots have remembered and forgotten. We then gather facts about the pilots, their past and present aviation experiences, the environment in which they operate, and some characteristics of the aeronautical knowledge itself. Finally, we attempt to link these characteristics of person, place, and thing to the observed patterns of pilot remembering and forgetting. Such an analysis should not only help identify knowledge areas that are prone to disuse and forgetting, but also help to reveal why.

In Experiment 1, we asked pilots to answer ten questions drawn from the FAA private pilot knowledge test to pilots and to provide us with details about themselves and their experiences. The data were analyzed to answer questions such as:

1. Does holding more certificates and ratings make pilots more likely to remember what they have learned?

2. Does more total or recent flight time lead to better retention?
3. Do flight instructors remember more after having taught the same material over and over again?
4. Do pilots remember more just before or after a flight review or practical test?

In Experiments 2 through 4, different pilots were asked to answer further questions to examine the influence of the kind of knowledge that pilots must remember, the materials that pilots use, and the characteristics of the places in which pilots fly. These data allow us to answer the following questions.

1. Is some aeronautical knowledge harder to remember than other knowedge?
2. Do the particular aircraft that pilots fly reinforce certain types of knowledge?
3. Do pilots better remember aeronautical knowledge that applies to their own familiar geographical regions?

We conclude by making some practical recommendations for how to improve the state of affairs for pilot knowledge retention, and by reviewing some important limitations of our study.

Experiment 1

In our first experiment, we administered ten questions drawn from the FAA private pilot knowledge exam to sixty pilots and asked them to provide us with details about seven aspects of their past and present aviation experience:

1. Certificates and ratings held;
2. Current role in aviation (active flight instructor, applicant for additional FAA certificate or rating, neither instructor nor applicant);
3. Total flight time;
4. Recent flight time (last 6 months, last 3 months);
5. Months since last flight review;
6. Months until next practical test (if applicant for additional certificate or rating);
7. Reading habits.

Apparatus

A paper and pencil, multiple-choice test was used for data collection. Each test contained the same ten questions randomly selected from the FAA private pilot item bank of questions. Questions that required extensive calculations (e.g., cross-country flight planning) were excluded, as were multiple questions drawn from the same topic area. The test was accompanied by a questionnaire that asked participants about the seven aspects of their past and recent aviation experience listed above.

Participants

Sixty pilots recruited from California Bay Area airports participated in the study. To insure a more uniform distribution of pilots across our seven aspects of pilot experience variables, we recruited pilots in equal numbers from the three categories of the current role in aviation variable. Twenty participants were active certified flight instructors (CFI group). Twenty participants held at least a private pilot certificate and were actively working toward an additional FAA pilot certificate or rating (Applicant group). Twenty participants were active pilots, holding at least a private pilot certificate, but were neither CFIs nor Applicants (Pilot group). Pilots were recruited for each category on a volunteer, first-come-first-served basis until each category was filled. Pilots received a NASA Aviation t-shirt as compensation for participating in the study.

Procedure

The experimental tests were completed by participants at times of their choosing. There was no time limit for completing the test. All participants were informed that their responses would remain anonymous.

Results and Discussion

Figure 1 shows a plot of the individual test scores for all pilots tested.

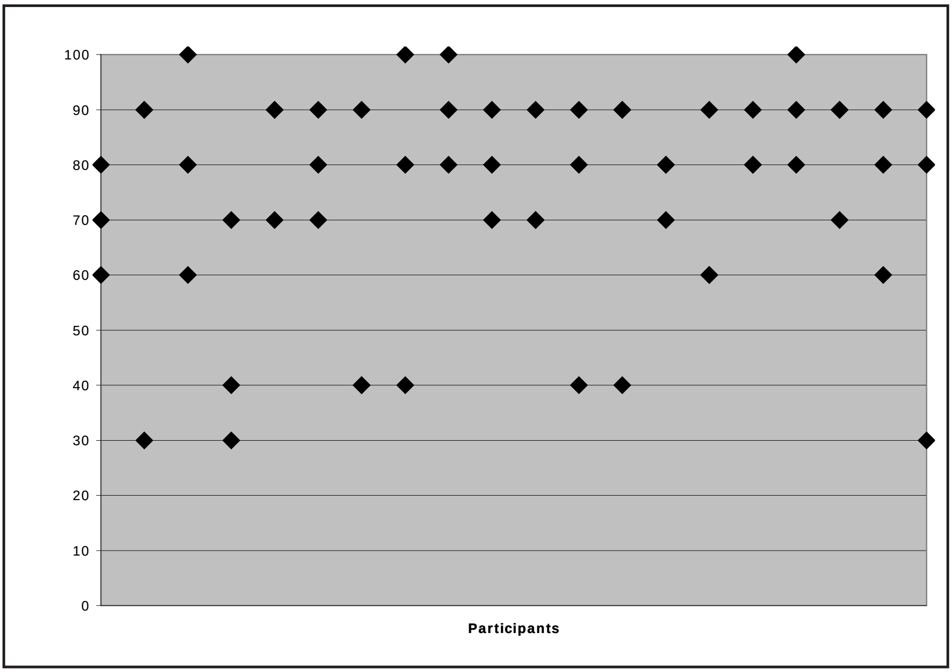


Figure 1. Test scores for all pilots tested.

Overall Test Performance. The results indicated a generally unimpressive overall performance. The average score for all sixty pilots was 74.8% with a standard deviation of 19.3%. Only 62% of all participants obtained a score higher than what is considered passing on the FAA private pilot knowledge test (70%). 15% of all participants obtained a score of 70%. 23% scored below 70%. Although a formal comparison was inappropriate due to the small sample size and limited variety of questions used here, it is interesting to note that only 38% scored higher than the national average score for the FAA private pilot airplane knowledge test (85%) (FAA, 2003).

It is important to reiterate that every participant in the study held at least an FAA private pilot certificate. That is, at some point in the past, every participant had achieved a passing score on the private pilot knowledge test from which the experimental test questions were drawn.

The data clearly showed that significant forgetting of the material tested by the FAA questions had taken place.

Certificates and Ratings Held. The scores for all pilots were segregated in four groups based on the certificates and ratings held by each pilot. The four groups and their mean scores were as follows: Private Pilot = 70.5% (21.1%); Private Pilot w/Instrument Rating = 77.8% (17.9%); Commercial Pilot = 72.2% (20.5%); and Certified Flight Instructor = 79.1% (17.6%). No significant differences were found among any of the four groups. Although there is considerable overlap in the aeronautical knowledge required for each successive pilot certificate, requiring pilots to study similar material repeatedly as they progress, the data did not indicate an improvement in retention due to training experience.

Current Role. The purpose of our three experimental groups was to measure the effect of the role that each pilot currently assumes on retention of aeronautical knowledge. It is important to note that this variable represents a notion different from that of certificates and ratings held by each pilot participant. The current role variable describes what each pilot is currently doing with the certificates and ratings that they hold. A participant in the Pilot group may have been a member of the Applicant group earlier that week before passing an Airline Transport Pilot practical test. Similarly, a member of the Applicant group may have been a member of the Pilot group a week earlier simply by deciding to pursue a Flight Instructor certificate. Thus, the three groups describe the status of pilot participants on the day and time that the test was administered.

The mean scores and standard deviations for the CFI, Applicant, and Pilot groups are shown in Table 1:

Table 1
Mean test scores and standard deviations for the three groups.

	CFI Group	**Applicant Group**	**Pilot Group**
Mean	79.0	76.0	69.5
Standard Deviation	18.0	17.0	22.1

A comparison of the means between the three groups revealed a marginally significant difference between the CFI and Pilot groups (df=18, t=1.49, $p < 0.09$). The large variability in scores among all three groups blurred the distinction between the means for all three groups. This result generally supports the idea that flight instructors rehearse their knowledge more often than other pilots do, and that leads to better retention. This result puts an interesting twist on the earlier finding about certificates and ratings held. Knowledge retention seemed to be affected not by the holding of certificates and ratings, but to some extent, what pilots are currently doing with those certificates and ratings.

Total and Recent Flight and Teaching Experience. The total and recent flight experience for all pilots tested is shown in Table 2, along with correlation coefficients comparing flight experience and scores on the experimental test.

Table 2
Correlations between test scores and total and recent flight experience.

	Total Flight Time		Past 6 Months		Past 3 Months	
	Hours	r	Hours	r	Hours	r
Pilot Group	382	.05	35	.21	13	.31
Applicant Group	272	.37	57	.14	32	-.21
CFI Group	1294	.04	178	**.47**	94	**.52**
All Three Groups	649	.11	90	**.34**	46	**.31**

There was little observed correlation between test scores and total flight experience. The three groups combined showed significant correlations between test scores and flight experience during the past six months (df=58, t=2.75, $p < .01$) and the past three months (df=58, t=2.48, $p < .01$). Most of this correlation is accounted for by the CFI group: past six months (df=18, t=2.26, $p < .05$), and past three months (df=18, t=2.58, $p < .01$).

Upcoming and Past Evaluations. There are generally two types of formal evaluations for the population of U.S. pilots: practical tests and flight reviews. The pilots in the Applicant group, by definition, were preparing for upcoming practical tests. All sixty of our pilot participants are subject to a flight review every 24 calendar months.

A significant negative correlation found was between test scores for the Applicant group and number of months remaining until the applicant's upcoming practical test (r=-0.68, df=18, t=3.93, $p < .005$). The closer each applicant was to their future practical test, the higher were their scores.

A similar correlation was found between test scores for the Pilot group and number of months since each pilot's last flight review ($r=-.44$, $df=18$, $t=1.96$, $p < .05$). Recently completed flight reviews were associated with higher scores. This result suggested that the flight review requirement only modestly serves to maintain pilot mastery of aeronautical knowledge.

Reading Habits. All pilots were asked to provide a Likert-type response to the question: "How often do you read magazines or books about flight training topics?" Interestingly, there were no differences in the reported reading frequency between the Pilot, Applicant, and CFI groups. Correlation coefficients for reading frequency and experimental test scores are shown in Table 3.

Table 3
Correlations between test scores and reported reading frequency

	r
Pilot Group	-.02
Applicant Group	.53 **
CFI Group	.05
All Three Groups	.14

A significant correlation was observed for the Applicant group ($df=18$, $t=2.65$, $p < .01$). The more time that these pilots reported that they spent reading, the better they did on the experimental test.

Summary

Overall, the seven aspects of pilot experience account for only modest portions of the variability in scores we observed. The data clearly showed that there is much more to the story about knowledge retention than certificates and ratings, flight time, and upcoming flight reviews and check rides. Pursuing these goals alone does not ensure that pilots will remember what they have learned.

Performance on individual questions

We then turned to a consideration of the ten test questions that pilots were asked to answer. This analysis allowed us to further explore links between which knowledge areas are used more frequently and which are retained and forgotten.

A breakdown of scores on each of the ten questions is shown in Figure 2. The results were further broken down by the current role for each pilot participant (i.e., CFI, Applicant, or Pilot).

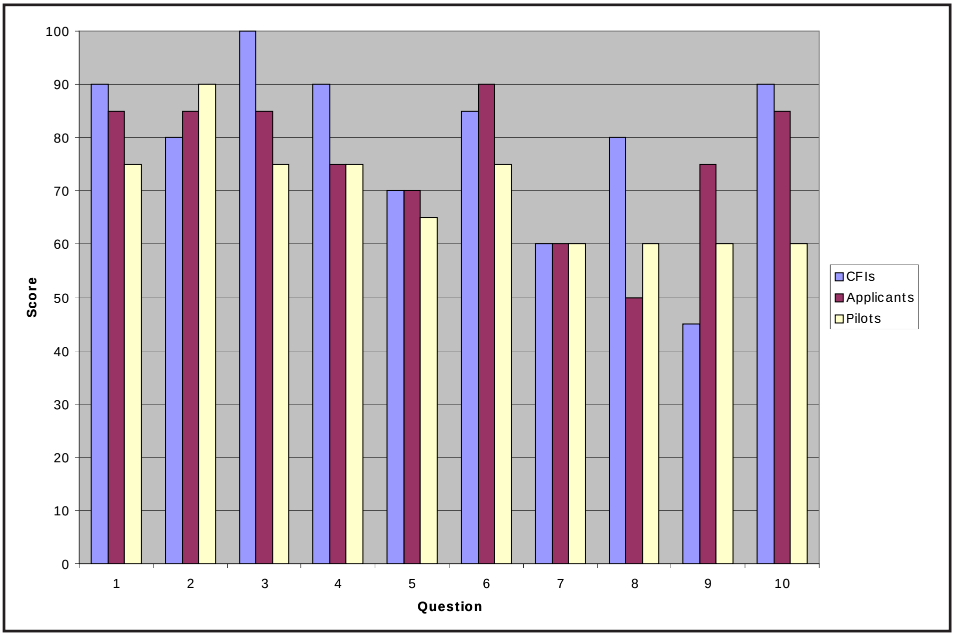


Figure 2. Scores for individual questions broken down by current role factor.

A correlation matrix comparing scores on individual questions to the remaining six factors of pilot experience is shown in Table 4. Significant correlations are shown in bold.

Table 4
Correlation coefficients comparing scores on individual questions to pilot experience.

	Question									
	1 Mag. Comp	2 W&B	3 Ceil-ing	4 Light Gun	5 Dens. Alt.	6 ELT	7 Night Flight	8 VOR	9 Alt. & Temp	10 Class G
Total Flight Time										
CFIs	-.07	.09	-	.13	.19	-.1	-.21	.33	-.29	.31
Applicants	.08	.05	.14	**-.44**	.2	-.7	.28	.27	.16	.11
Pilots	.03	.18	.3	.18	.17	**-.4**	-.16	-.2	-.02	.17
Combined	.06	.03	.2	-.01	.17	-.2	-.02	**.27**	.17	**.24**
Last 6 Months										
CFIs	.31	.16	-	.31	0.0	.2	.24	**.46**	.17	.33
Applicants	.29	-.15	.27	**-.56**	.15	0.0	.34	.17	.06	0.0
Pilots	.2	.05	.24	.24	.17	.2	.23	.23	**-.39**	-.12
Combined	**.25**	.02	.2	.17	.05	.2	.17	**.34**	-.06	**.22**
Last 3 Months										

CFIs	**.4**	.17	-	.37	-.05	.3	.2	**.58**	.27	.18
Applicants	.24	0.0	-.09	**-.61**	-.11	0.0	.13	-.01	-.22	.11
Pilots	.17	-.15	.25	.2	.22	.1	.24	.31	-.31	**.39**
Combined	**.28**	.01	.19	.13	0.0	.2	.12	**.34**	-.04	**.25**
Reading Frequency										
CFIs	-.35	-.2	-	-.13	-.04	**.6**	.08	0.0	.16	-.02
Applicants	.12	-.22	**.81**	.14	.37	-.2	.3	.2	.33	.13
Pilots	.05	.26	.25	.12	.2	-.3	-.19	-.25	-.36	.16
Combined	-.03	-.05	**.35**	.05	.17	0	.04	-.03	.01	.11
Next Practical										
Applicants	-.1	.17	**-.7**	-.21	**-.5**	.1	**-.57**	-.2	**-.46**	-.1
Last Flight Review										
Pilots	.21	.11	-.33	**-.5**	-.36	**-.4**	-.24	-.2	.07	-.2
Certificates/Ratings										
Applicants	-.02	-.02	-.02	-.07	-.15	.2	.37	-.1	-.25	.15
Pilots	.26	.33	.26	.11	-.1	0.0	0.0	0.0	.13	-.13
Combined	.12	.14	.12	.02	-.13	0	.2	0	-.07	-.03

The analysis of the ten individual questions below makes frequent reference to the data presented in Figure 2 and Table 4.

Question 1. In the Northern Hemisphere, a magnetic compass will normally indicate initially a turn toward the east if:

A—an aircraft is decelerated while on a south heading.
B—an aircraft is accelerated while on a north heading.
C—a left turn is entered from a north heading.

This question asked about a compass turning error that is likely observed more in a training environment, and only occasionally during training, since the magnetic compass is not considered a primary flight instrument. Indeed, the significant correlations between scores on this question and flight experience during the past three months were mainly due to CFIs and applicants. The Pilot group also performed admirably on this question, perhaps because pilots are taught a memory aid to help them to remember this turning error (ANDS – Accelerate North, Decelerate South) (FAA, 1999; Jeppesen Sanderson, 1999; Kershner, 2001; Gardner, 2002). The effectiveness of mnemonics like this one has been widely demonstrated (Baddeley, 1998; Yates, 1966).

Question 2. (Refer to Figure 3) Determine if the airplane weight and balance is within limits.

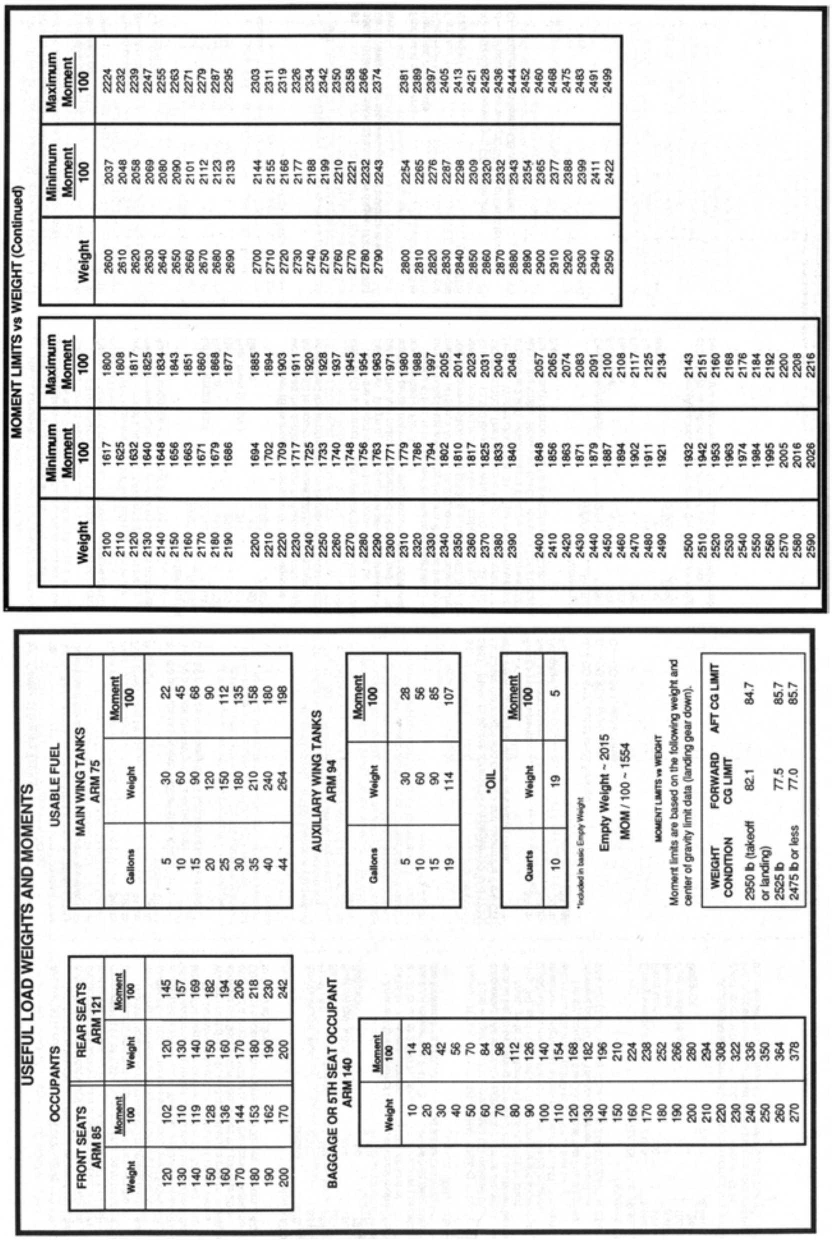
Front seat occupants……………………………..415 lb
Rear seat occupants………………………………110lb

Fuel, main tanks……………………………….44 gal
Fuel, aux. Tanks………………………………..19 gal
Baggage……………………………………….32 lb

A—19 pounds overweight, CG within limits.

B—19 pounds overweight, CG out of limits forward.

C—Weight within limits, CG out of limits.



USEFUL LOAD WEIGHTS AND MOMENTS

OCCUPANTS

FRONT SEATS ARM 85		REAR SEATS ARM 121	
Weight	Moment/100	Weight	Moment/100
120	102	120	145
130	110	130	157
140	119	140	169
150	128	150	182
160	136	160	194
170	144	170	206
180	153	180	218
190	162	190	230
200	170	200	242

BAGGAGE OR 5TH SEAT OCCUPANT ARM 140

Weight	Moment/100
10	14
20	28
30	42
40	56
50	70
60	84
70	98
80	112
90	126
100	140
110	154
120	168
130	182
140	196
150	210
160	224
170	238
180	252
190	266
200	280
210	294
220	308
230	322
240	336
250	350
260	364
270	378

USABLE FUEL

MAIN WING TANKS ARM 75

Gallons	Weight	Moment/100
5	30	22
10	60	45
15	90	68
20	120	90
25	150	112
30	180	135
35	210	158
40	240	180
44	264	198

AUXILIARY WING TANKS ARM 94

Gallons	Weight	Moment/100
5	30	28
10	60	56
15	90	85
19	114	107

*OIL

Quarts	Weight	Moment/100
10	19	5

*Included in basic Empty Weight

Empty Weight ~ 2015
MOM / 100 ~ 1554

MOMENT LIMITS vs WEIGHT

Moment limits are based on the following weight and center of gravity limit data (landing gear down).

WEIGHT CONDITION	FORWARD CG LIMIT	AFT CG LIMIT
2950 lb (takeoff or landing)	82.1	84.7
2525 lb	77.5	85.7
2475 lb or less	77.0	85.7

MOMENT LIMITS vs WEIGHT (Continued)

Weight	Minimum Moment/100	Maximum Moment/100
2100	1617	1800
2110	1625	1808
2120	1632	1817
2130	1640	1825
2140	1648	1834
2150	1656	1843
2160	1663	1851
2170	1671	1860
2180	1679	1868
2190	1686	1877
2200	1694	1885
2210	1702	1894
2220	1709	1903
2230	1717	1911
2240	1725	1920
2250	1733	1928
2260	1740	1937
2270	1748	1945
2280	1756	1954
2290	1763	1963
2300	1771	1971
2310	1779	1980
2320	1786	1988
2330	1794	1997
2340	1802	2005
2350	1810	2014
2360	1817	2023
2370	1825	2031
2380	1833	2040
2390	1840	2048
2400	1848	2057
2410	1856	2065
2420	1863	2074
2430	1871	2083
2440	1879	2091
2450	1887	2100
2460	1894	2108
2470	1902	2117
2480	1911	2125
2490	1921	2134
2500	1932	2143
2510	1942	2151
2520	1953	2160
2530	1963	2168
2540	1974	2176
2550	1984	2184
2560	1995	2192
2570	2005	2200
2580	2016	2208
2590	2026	2216

Weight	Minimum Moment/100	Maximum Moment/100
2600	2037	2224
2610	2048	2232
2620	2058	2239
2630	2069	2247
2640	2080	2255
2650	2090	2263
2660	2101	2271
2670	2112	2279
2680	2123	2287
2690	2133	2295
2700	2144	2303
2710	2155	2311
2720	2166	2319
2730	2177	2326
2740	2188	2334
2750	2199	2342
2760	2210	2350
2770	2221	2358
2780	2232	2366
2790	2243	2374
2800	2254	2381
2810	2265	2389
2820	2276	2397
2830	2287	2405
2840	2298	2413
2850	2309	2421
2860	2320	2428
2870	2332	2436
2880	2343	2444
2890	2354	2452
2900	2365	2460
2910	2377	2468
2920	2388	2475
2930	2399	2483
2940	2411	2491
2950	2422	2499

Figure 3. Weight and balance charts for Question 2

The correct answer to this question is that the aircraft is not approved to fly as it is currently loaded. The Pilot group performed best while only eighty percent of the CFI group answered the question correctly. This might be attributed to flight instructors not working many weight and balance problems. Instructors might leave this type of problem solving to students to learn on their own, or in a ground school. In addition, training flights seldom exceed the weight and balance limitations and instructors might simply skip the calculations. Pilots flying outside of a training environment may be more often confronted with novel aircraft loading situations. There were no significant correlations between scores on this question and any of the seven factors we considered.

Another possible explanation for less than perfect performance is a well-known outcome of the learning process (Anderson, 2000). After initially learning the concepts needed to solve weight and balance problems, pilots' skills may become tuned to the particular charts they currently use. Their ability to solve weight and balance problems with unfamiliar charts may be less, until they have had a chance to practice a few times with them. Indeed, every insurance company requires every pilot, regardless of experience level, to complete a checkout for each new make and model of aircraft they intend to fly.

Question 3. For aviation purposes, ceiling is defined as the height above the Earth's surface of the

A – lowest reported obscuration and the highest layer of clouds reported as overcast.
B – lowest broken or overcast layer or vertical visibility into an obscuration.
C – lowest layer of clouds reported as scattered, broken, or thin.

The three groups of pilots collectively recorded the highest average score on this question. Information about cloud layers and obscurations is given in METARs. The term ceiling is used when describing cloud layers and obscurations in ATIS, AWOS, and ASOS reports. Hence, it is reasonable to assume that all pilots get regular practice with dealing with the definition of a ceiling. There was a strong correlation between scores on this question and reading frequency among the Applicant group.

Question 4. While on final approach for landing, an alternating green and red light followed by a flashing red light is received from the control tower. Under these circumstances, the pilot should

A—discontinue the approach, fly the same traffic pattern and approach again, and land.
B—exercise extreme caution and abandon the approach, realizing the airport is unsafe for landing.
C—abandon the approach, circle the airport to the right, and expect a flashing white light when the airport is safe for landing.

Question 4 is a classic example of an emergency procedure that is seldom rehearsed outside of flight training environments. However, light gun signals, to some extent, make use of universal conventions, namely that green generally signals that it is okay to proceed, while red signals stop. These conventions are well known to aid memory (Yates, 1966). Interestingly, there were strong negative correlations between scores for this question and total and recent flight experience for the Applicant group.

Question 5. (Refer to Figure 4) Determine the density altitude for these conditions:

Altimeter setting …………………………………….. 30.35
Runway temperature ……………………………… + 25 F
Airport elevation …………………..…….. 3,894 ft MSL

A—2,000 feet MSL.
B—2,900 feet MSL.
C—3,500 feet MSL.

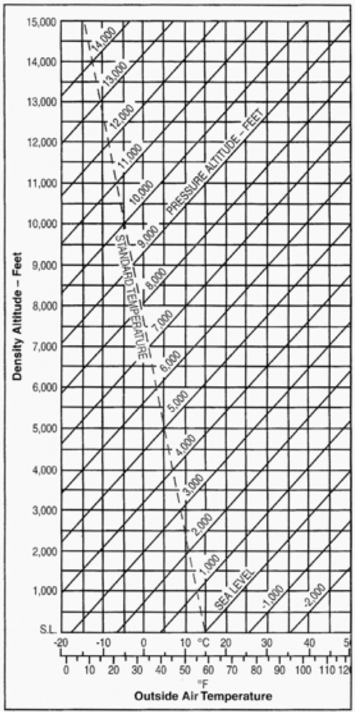
Question 5 is a simple density altitude problem. Only 70 percent of participants in the CFI and Applicant groups, and 65 percent in the Pilot group answered the question correctly. This is particularly worrisome given that density altitude calculation is an important, often critical, skill. One possible explanation for this result is that all participants in the study were drawn from the coastal region of California where high density altitude is seldom encountered. Another possible explanation is that many aircraft performance charts integrate density altitude information into the chart, eliminating the need to calculate density altitude explicitly. Hence, that technique may have fallen to disuse. There was a strong correlation between scores on this question and time remaining until an upcoming practical test for pilots in the Applicant group: the closer they got to the practical test, the better they did on the density altitude question.

Question 6. When are non-rechargeable batteries of an emergency locator transmitter (ELT) required to be replaced?

A—Every 24 months.
B—When 50 percent of their useful life expires.
C—At the time of each 100-hour or annual inspection.

The Applicant group recorded the best performance on this question, and good performance was strongly associated with greater reading frequency. A possible explanation for the moderate performance among the CFI and Pilot groups might be that many flying schools/clubs have a maintenance board showing due dates for required inspections. Or that club managers ensure that airplanes meet inspection requirements. This knowledge is also supported by a mnemonic. ELT batteries must be replaced every “12-1-1/2”: every 12 calendar months, after 1 hour of continuous use, or after 1/2 of their useful life has expired.

DENSITY ALTITUDE CHART



Altimeter Setting (" Hg)	Pressure Altitude Conversion Factor
28.0	1,824
28.1	1,727
28.2	1,630
28.3	1,533
28.4	1,436
28.5	1,340
28.6	1,244
28.7	1,148
28.8	1,053
28.9	957
29.0	863
29.1	768
29.2	673
29.3	579
29.4	485
29.5	392
29.6	298
29.7	205
29.8	112
29.9	20
29.92	0
30.0	-73
30.1	-165
30.2	-257
30.3	-348
30.4	-440
30.5	-531
30.6	-622
30.7	-712
30.8	-803
30.9	-893
31.0	-983

Figure 4. Density altitude chart for Question 5.

Question 7. If recency of experience requirements for night flight are not met and official sunset is 1830, the latest time passengers may be carried is

A – 1929.
B – 1829.
C – 1859.

Question 7 asked pilots to determine how late in the day they can legally fly with passengers when they are not "night current." It is reasonable to assume that decisions about night currency come up frequently in everyday practice. Surprisingly, forty percent of pilots in all three groups answered this question incorrectly. This regulation is another example of a fact that must be memorized, in this case, without the benefit of a memory aid or mnemonic. A further explanation for the poor performance is that pilots may use their own informal rules for estimating the beginning of official night. Being on the ground well before darkness would allow pilots to easily abide by the regulation without remembering its specifics. The only encouraging result was pilots in the Applicant group did better on this question as their upcoming practical test drew near.

Question 8. (Refer to Figure 5) The VOR receiver has the indications shown. What is the aircraft's position relative to the station?

A – North.
B – East.
C – South.

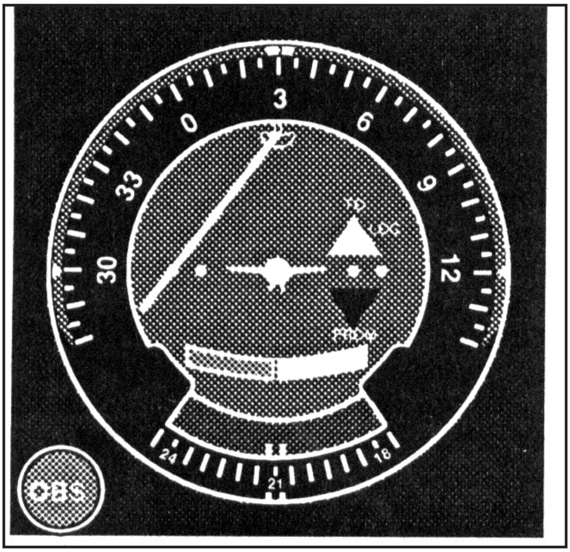


Figure 5. VOR illustration for Question 8

Question 8 asked pilots to determine their position with respect to a VOR from a single VOR indication. CFIs were the only pilots to record an even acceptable performance on this question, and this success was significantly correlated with recent flight experience. Surprisingly, only 50% of the Applicant group answered this question correctly. The most common additional rating sought by applicants is the instrument rating. VOR interpretation, of course, is a fundamental instrument skill. One explanation is that the question asks pilots to interpret a CDI indication in a way that is different from how a VOR is typically used. When determining one's position with respect to a VOR, most pilots turn the OBS knob until the needle is centered with a FROM indication, and note that they are positioned along the indicated radial. At some point, they may have learned the technique required by the question but that technique has fallen to disuse. Another possible explanation for the Pilot group is that VOR skills may be used less as the popularity of GPS navigation increases.

Question 9. How do variations in temperature affect the altimeter?

A—Pressure levels are raised on warm days and the indicated altitude is l ower than true altitude.

B—Higher temperatures expand the pressure levels and the indicated altitude is higher than true altitude.
C—Lower temperatures lower the pressure levels and the indicated altitude is lower than true altitude.

Question 9 asked pilots to apply an informal rule about changing pressure and temperature: “From high to low, look out below.” Average scores for the CFI and Pilot groups were surprisingly low for this question: 45% and 60%, respectively. One explanation for the poor scores is that there is no one accepted procedure for dealing with the effect of cold temperatures on indicated altitude. Pilots might decide to climb a few hundred feet above their planned or assigned altitude, but would have to do so at their own risk. In the absence of specific guidance, pilot may choose to disregard temperature effects. The Applicant group performed acceptably well on this question and their success was significantly correlated with time to upcoming practical test.

Question 10. What minimum visibility and clearance from clouds are required for VFR operations in Class G airspace at 700 feet AGL or below during daylight hours

A – 1 mile visibility and clear of clouds.
B – 1 mile visibility, 500 feet below, 1,000 feet above, and 2,000 feet horizontal.
C – 3 miles visibility and clear of clouds.

Question 10 asked pilots to recall a fact about VFR weather minimums in Class G airspace. The CFI and Applicant groups scored admirably well (90% and 85%, respectively). The Pilot group performed poorly (60%) and their limited successes seemed to be associated with higher flight times within the past three months. Low scores in the Pilot group may reflect the absence of an easy-to-remember memory aid for Class G minimums, or be an example of how experienced pilots often simplify difficult knowledge. The weather minimums for Question 10 are the most liberal of any type of airspace (1 SM, and clear of clouds). Pilots may choose not to rehearse this knowledge, adopting a more conservative weather minimum that allows them to abide by the regulation. It is likely a rare occasion when a pilot takes off in 1 SM and clear with the intention of remaining within 700 feet AGL. A more typical scenario is to climb to at least the traffic pattern altitude, or depart the traffic pattern, enter Class E airspace, and abide by more conservative and easy to remember weather minimums.

Summary

The analysis of individual questions casts considerable doubt on simple hypotheses about what kinds of knowledge are used and remembered and what kinds are not. The analysis proposed several more subtle factors that might help explain what is remembered and what is forgotten.

Mnemonics

Several of the questions appearing on the knowledge tested required pilots to recall facts. Some of these facts (Questions 1 and 6) were supported by popular memory aids or mnemonics that appear routinely in the textbooks used by private pilot students, while others were not (Questions 7 and 10). A comparison of scores between these two groupings of questions yielded a significant difference ($df=5$, $t=2.19$, $p < 0.05$), supporting the already widely recognized usefulness of mnemonics.

Specialization

A well-known outcome of the learning process suggested by the weight and balance and density altitude questions (Questions 2 and 5) was that as people acquire skill, their knowledge and methods tend to become finely tuned to the particular procedures and materials they use, while more general knowledge and skill becomes less available (Greeno, 1974; Logan, 1988). Pilots' ability to work problems such as weight and balance and density altitude may be highest when using familiar charts, but less when using different charts.

Simplification

A striking characteristic of many aviation regulations is that they contain intricate and sometimes similar details. The weather minimums for Class G airspace, tested by Question 10, require the pilot to remember minimum visibilities and cloud clearances for five different cases. The rules for carrying passengers at night, tested by Question 7, contain details that are similar but slightly different from related rules that govern the use of aircraft lights and the logging of night flight. The potential for memory decay and interference (Anderson, 2000) is widely recognized. A well-known outcome of the learning process is that, as people acquire expertise in a domain, they seek shortcuts and simplifications for difficult problems (Blessing & Anderson, 1996; Koedinger & Anderson, 1990; Casner & Larkin, 1989). Pilots might simplify these rules by adopting a higher standard that allows them to remain legal in all cases, and excuses them from having to remember the details.

Characteristics of the Pilot's Geographical Environment

The analysis of questions 5 and 9 proposed the idea that different geographical areas may afford opportunities to rehearse and retain, or disuse and forget, different aeronautical knowledge. Our sample of pilots was drawn from the coastal region of California. In their home environs, these pilots seldom encounter high density altitudes or extremely cold temperatures. Furthermore, this reduced opportunity to rehearse these concepts may have been accompanied by a reduced emphasis on the same knowledge areas during training. These factors might explain the modest performance we observed on questions 5 and 9. Perhaps pilots from a geographical area in which high elevations and extreme temperatures are common might do better.

Experiment 2

The purpose of this experiment was to test our hypothesis about knowledge specialization: that pilots' knowledge may become fine-tuned to the charts and procedures associated with aircraft they fly, and that more general problem-solving knowledge they learned during initial training gradually fades.

To test this hypothesis, we recruited a sample of pilots who flew regularly in one make and model single-engine airplane and who had never flown in a second make and model single-engine airplane. These pilots were asked to solve weight and balance problems in both airplanes. It is important to note that all pilot participants held at least a private pilot certificate with an airplane category and single-engine class rating. This means that all pilots were certified to operate any (non-turbojet) single-engine airplane.

If our hypothesis about knowledge specialization is correct, pilots will be more successful at solving the weight and balance problems in the familiar airplane.

Apparatus

A paper and pencil test was used for data collection. Each test contained three weight and balance problems drawn from a test bank of four possible problems as follows. Two problems used weight and balance charts for a single-engine domestic airplane for which all pilots had significant experience and had flown within the preceding days. The remaining two problems used weight and balance charts for a different single engine domestic airplane that none of the pilots had ever flown. One problem for each manufacturer's charts resulted in a within-limits solution, while the other problem resulted in an out-of-limits or "no go" solution. Each problem required pilots to do three things: (1) calculate gross weight; (2) calculate total moments; and (3) determine whether the airplane was safe to fly as loaded. The test was accompanied by a questionnaire that asked participants about the certificates and ratings they hold and their total and recent flight time.

Participants

Twenty-four current and active pilots recruited from local California Bay Area airports participated in the study. Pilots received a NASA Aviation t-shirt as compensation for participating in the study.

Procedure

The experimental tests were completed by participants at times of their choosing. There was no time limit for completing the test. All participants were informed that their responses would remain anonymous.

Results and Discussion

The results for the four problems are shown in Table 5. Each problem was graded using three criteria: (1) correct weight calculation; (2) correct balance calculation; and (3) correct decision about whether the airplane was loaded within limits.

Table 5
Mean scores for weight and balance problems

Familiar Airplane						Unfamiliar Airplane					
Within Limits			Out-Of-Limits			Within Limits			Out-Of-Limits		
Wt.	Bal.	Go?	Wt.	Bal.	Go?	Wt.	Bal.	Go?	Wt.	Bal.	Go?
.83	.89	.78	1.0	.94	.83	.94	.83	.78	.94	.5	.5

There was a significant difference between the Unfamiliar Airplane Out-Of-Limits problem and all other problems. No other significant differences between the other problems were found.

Table 6 shows correlation coefficients comparing scores on the weight and balance problems and total and recent flight and teaching experience.

Table 6
Correlations between recent flight and teaching experience and scores on weight and balance problems.

Total Flight Time (n=18)	Last 6 Months n=18)	Total Dual Given (n=11)	Dual Last 6 Months (n=11)
-.21	.19	-.65	.28

A slight negative correlation was observed between total flight time and test scores. This effect was reversed in the presence of higher recent flight experience.

A significant negative correlation was observed between test scores and total dual instruction given among the flight instructor participants ($df=9$, $t=2.26$, $p < .05$). This effect was reversed in the presence of higher recent dual instruction given experience.

The results suggested two interesting conclusions. First, there is reasonable evidence to support the hypothesis that pilots well retain the particulars of the aircraft weight and balance charts they use everyday, and are less skilled at using charts for difference airplanes. Pilot who had never flown our control airplane were able to recognize a “no go” situation only 50 percent of the time. It can be argued that this result is natural: problem solving will be better for anyone using familiar materials. However, it must be remembered that holding a pilot certificate with a category and class rating means that the pilot is privileged to operate any aircraft of that category and class – without any further formal training or evaluation.

Second, there is little reason to believe that total flight time means higher proficiency in solving weight and balance problems. Furthermore, it seems that as flight instructors spend more time in the dual instruction role, the worse they become at solving weight and balance problems. This effect seems to be mitigated as the flight instructor's recent teaching experience increases.

Lastly, we must be careful in comparing the scores on weight and balance problems for this experiment, and the scores observed on the weight and balance problem in Experiment 1. The problem used in Experiment 1 was a multiple-choice FAA test question (1-in-3 chance of a successful guess), whereas the problems used in this experiment graded pilots across each step of their work.

Experiment 3

The purpose of this experiment was to test the hypothesis that pilots develop and use simplifications for aeronautical knowledge that requires tedious rote memorization. In the case of regulations, a simplification might discard difficult-to-remember details in favor of a simpler rule that, while not correct according to the regulations, allows pilots to operate legally.

To illustrate the notion of simplification, suppose a pilot states that, for all Class G airspace situations, the minimum visibility is 5 statute miles, while the minimum distance from clouds is 1,000 ft. above and below, and 1 statute mile horizontal. This simplification results in knowledge that is incorrect according to the regulations, yet allows him to operate legally in all Class G airspace situations. Now suppose a different pilot states that the minimums for all Class G airspace are 1 statute mile visibility and clear of clouds. This simplification results in knowledge that is both incorrect and not legal.

Our distinction between correct and legal answers affords us the opportunity to explore one more interesting twist: how certain pilots are about their answers. That is, if pilots are using simplifications, are they aware of them? Will pilots who provide "merely legal" answers recognize this situation, or will they confidently (and mistakenly) say that their answers are correct and "by the book?"

To answer these questions, we asked a group of pilots to answer six questions about regulations, and scored their answers as correct, legal, or altogether wrong. For each question, pilots were also asked to indicate (yes or no) if they were certain that their answer was correct according to the regulations, or if they were uncertain and would use their answer to operate legally.

If the hypothesis about knowledge simplification is correct, pilots will provide answers that are legal but not necessarily correct.

Apparatus

A paper and pencil, multiple-choice was used for data collection. Each test contained the same six questions, in shuffled order.

Three questions asked pilots to supply VFR weather minimums for Class G airspace in three different situations (14 CFR 91.155):

- Day, 1,200 ft. or less;
- Day, more than 1,200 ft. but less than 10,000 ft;
- Night 1,200 ft. or less.

The three remaining questions asked pilots about rules for operating at night:

- What time can a pilot begin logging night flight (14 CFR 1.1)?
- At what time must an airplane have operational position lights (14 CFR 91.209)?
- What time must passengers be dropped off if the pilot has not met the recent flight experience requirements for night flight (14 CFR 61.57(b))?

It is important to note that none of the questions were trick questions or an attempt to "split hairs" by offering answer choices that differed by one minute in order to test pilots' understanding of the minutiae of a rule. A basic understanding sufficed to answer all questions.

Participants

Eighteen current and active pilots recruited from local California Bay Area airports participated in the study. Pilots received a NASA Aviation t-shirt as compensation for participating in the study.

Procedure

The experimental tests were completed by participants at times of their choosing. There was no time limit for completing the test. All participants were informed that their responses would remain anonymous.

Results and Discussion

The results for the six questions are shown in Table 7.

Table 7
Scores and certainty measures for regulations questions.

Class G 1			**Class G 2**			**Class G 3**		
Correct	Legal	Wrong	Correct	Legal	Wrong	Correct	Legal	Wrong
.67	1.0	0	.5	1.0	0	.89	1.0	0
Certainty								
.72			.67			.78		
Correlation: Certainty / Correctness								
r=.35			**r=.47**			r=-.19		

Night 1			Night 2			Night 3		
Correct	Legal	Wrong	Correct	Legal	Wrong	Correct	Legal	Wrong
.61	.72	.28	.56	.89	.11	.44	.89	.11
Certainty								
.78			.83			.72		
Correlation: Certainty / Correctness								
r=.12			r=.06			r=.31		
Correlation: Legal / Correctness								
r=-.03			r=-.16			r=-.22		

The results directly supported the hypothesis about knowledge simplification: pilots characteristically gave incorrect yet legal responses.

Perhaps the most interesting results pertain to the certainty measures. Despite being given the option to say they were unsure, pilots frequently stated that they had provided the correct answer when in fact they had provided a merely legal answer, or an answer that was neither correct nor legal. On only one question did pilots' certainty significantly correlate with the correctness of their responses. This suggested that pilots had not only forgotten the regulations but were also unaware they had forgotten them. The certainty data also ruled out the theory that pilots offload the burden of remembering weather minimums and simply look them up prior to flight, or rely on notes during flight. If pilots followed such a strategy, it seems unlikely that their certainty estimates would be so high and so far amiss.

Experiment 4

This last experiment aimed to test our hypothesis that pilots in different geographical areas would exhibit greater retention of aeronautical knowledge that was more applicable to their own environment. To test this hypothesis, we selected eight questions about density altitude and airplane performance from the FAA Private Pilot knowledge test bank and administered them to pilots in two geographical areas: (1) the California coast during the winter; and (2) Denver, Colorado during the summer. Four questions were "conceptual questions" that probed pilots' understanding of the concepts that underlie density altitude. The remaining four questions asked pilots to use charts, perform calculations, and solve density altitude and airplane performance problems. The average elevation of the California airports at which pilots were recruited was 28 ft. The average elevation of the Colorado airports was 5770 ft. The average daily peak temperature in California during data collection was approximately 50 degrees Fahrenheit. The average daily peak temperature in Colorado during data collection was approximately 95 degrees Fahrenheit.

Apparatus

A paper and pencil, multiple-choice test was used for data collection. Each test contained the above-described eight density altitude questions. The test was accompanied by a questionnaire that asked participants about the certificates and rating they held and their total and recent flight time.

Participants

Thirty-six current and active pilots participated in the experiment. Eighteen pilots were recruited from local California Bay Area airports during winter. The remaining eighteen pilots were recruited at two airports located in Denver, Colorado during the summer. The mean and standard deviation for total number of flight hours for pilots in the California group and Colorado group were 1025 (978) and 936 (1238), respectively. Pilots received a NASA Aviation t-shirt as compensation for participating in the study.

Procedure

The experimental tests were completed by participants at times of their choosing. There was no time limit for completing the test. All participants were informed that their responses would remain anonymous.

Results and Discussion

The mean scores and standard deviations for the two groups are shown in Table 8.

Table 8
Mean scores for density altitude test for California and Colorado groups.

	Overall Score	Concept Questions	Problems
California Coast	.85 (.13)	.97 (.24)	.74 (.08)
Denver, Colorado	.86 (.14)	.97 (.23)	.75 (.08)

The scores for the two groups were nearly identical, offering no support for our hypothesis that pilots who operate everyday in high density altitude conditions know more than pilots who operate at sea level in a cool climate. This result is both surprising and counterintuitive. There are a number of possible explanations for this outcome, and for why the hypothesis may warrant further investigation.

First, most pilot participants in both groups were students and flight instructors who worked in a training environment at local flight schools. It may be that these two environments are more similar than we suspected. The airplanes used at each flight school were able to take off, climb, and land at any time of the day at either location. Furthermore, there is no significant terrain within close proximity of either airport to make climb rates an immediate safety issue. A future study that recruited workaday pilots who fly between small mountain airports might find that these pilots exercise their knowledge about density altitude and airplane performance to a larger extent, and therefore retain a greater mastery.

Second, our third experiment established that pilots devise and use simplifications of aeronautical knowledge they have learned. There are a number of "rules of thumb" that can be used in lieu of performing more tedious density altitude and takeoff performance calculations. For example, at an average field elevation of 5,770 feet, density altitude can be approximated by simply adding two zeros to the temperature in Fahrenheit. Depending on atmospheric pressure, density altitude is roughly 7,000 feet at 70 degrees, 8,000 feet at 80 degrees, 9,000 feet at 90 degrees, etc. Pilots may also rely on practical rules for takeoff performance such as the "70-50" rule: if the airplane has not developed 70% of the target rotation speed after using 50% of the available runway, the takeoff should be aborted.

Lastly, it may be that our decision to use FAA test questions to test what pilots know about density altitude and airplane performance was entirely insensitive to what knowledge pilots retained, and what new knowledge they have acquired. Perhaps a future study that undertook a more detailed review of pilot knowledge, beyond standardized multiple-choice questions, could reveal differences in what pilots know.

Summary and Conclusions

Four experiments were conducted to help understand how well everyday flying experience provides pilots with the opportunity to rehearse and retain aeronautical knowledge. In our first experiment, ten questions selected from the FAA private pilot knowledge test were given to a group of sixty pilots. The average score for all ten questions was 74.8%. Using information we gathered from the sixty pilots about their past and present aviation activities, we attempted to relate retention of the aeronautical knowledge tested by the ten questions to seven aspects of pilot experience. It was found that the certificates and ratings held by pilots have little influence on how well those pilots retain what they have learned during training. The current role played by the pilot (teacher, student, or neither) has a marginally significant effect: better retention is indicated for pilots who currently teach or learn in a flight training environment. Total flight experience seems to have little effect on knowledge retention while recent flight experience is associated with better retention, especially for flight instructors. Proximity in time to past flight reviews and upcoming practical tests was associated with better retention. Finally, better retention was found among applicants who reported frequently reading more aviation-related materials. Overall, the seven factors of pilot experience we considered accounted for portions of the variability in test scores, but left much unexplained.

An analysis of the ten individual test questions led us to make three additional hypotheses, not about pilots, but about the characteristics of the aeronautical knowledge itself, the charts that pilots use, and the geographical regions in which pilots use their knowledge. It was shown how aeronautical knowledge that requires rote memorization of facts and details is better remembered when mnemonics are used. Our second experiment showed how pilots had become accustomed to the weight and balance charts found in airplanes that are familiar to them, while their ability to solve the same weight and balance problems in other airplanes was less than what is expected of a private pilot applicant. The third experiment demonstrated how pilots appear to substitute simplified rules in place of more complex regulations that require memorization of detailed facts, and how pilots are often unaware that they have made this simplification. A fourth experiment examined the effect of geographical region on retention of aeronautical knowledge by testing pilots for superior retention of knowledge that pertained to their own local weather patterns. It was found that pilots who taught and trained in a high-density altitude environment performed no better or worse than pilots who taught and trained in a cool, coastal climate.

Overall, the results cast considerable doubt on the assumption that everyday flying or teaching experience, together with the current recent flight experience and flight review requirements, will naturally offer pilots the opportunity to practice and keep fresh the entirety of what they have learned. The results indicate a need for regular study, not only in areas of suspected disuse, such as regulations, emergencies, and unfamiliar weather patterns, but also in areas that seem more familiar. Indeed, the results disconfirm simple theories about what knowledge is used regularly and retained and what knowledge is not. The results for weight and balance problems using familiar aircraft charts demonstrate that pilots may not get as much practice in some areas as our intuitions may suggest. The certainty measures associated with incorrect responses to questions about regulations further demonstrate that pilots do not always know what they do not know.

The results indicated a need for more explicit standards for ongoing aeronautical knowledge proficiency. One possibility might be a system of self-certification similar to the recent flight experience requirements specified by U.S. 14 CFR 61.57. Pilots complete the requisite number of takeoffs, landings, or instrument procedures, and note these events in their pilot logbooks. Another, perhaps more controversial, alternative is to create a more detailed and frequent evaluation for aeronautical knowledge beyond the one-hour flight review required by U.S. 14 CFR 61.56.

Limitations of Our Study

There are a number of limitations of our study that warrant consideration.

What Do Pilots Really Need To Know?

A key assumption of our study was that pilots needed to know the aeronautical knowledge tested by the questions drawn from the FAA private pilot knowledge test bank. It might be argued that, even though the knowledge is required for the FAA knowledge test, the material covered by these questions did not adequately represent what a pilot operating in the national airspace system really needs to know. This argument seems potentially valid for some of our original ten test questions. For example, our own hypothesis could be used to argue that what is important is that pilots be able to solve weight and balance problems in their own airplanes. Pilots' ability to solve these problems using the different charts found in FAA questions may not matter much. However, this argument is rejected by the modest scores achieved by pilots in our second experiment in which familiar weight and balance charts were used. Making a correct go/no-go decision four out of five times in a familiar aircraft is simply not sufficient. Another argument could be made about the question in which pilots were asked to interpret a VOR indication without being able to center the needle. A convincing argument would need to prove that no pilot could ever look at a VOR indication configured as such and draw an erroneous conclusion that subsequently went uncorrected.

What Do Other Kinds of Pilots Remember?

Another limitation of our study is that we only considered pilots drawn from a small portion of the general aviation population: pilots who teach, learn, and fly at flight training/aircraft rental establishments in metropolitan areas. We have already recognized the possibility that pilots who fly in other areas, under less routine circumstances in different geographical areas, may exhibit different knowledge retention than what we have observed. Perhaps airline, medical evacuation, and crop-duster pilots all exhibit different patterns of remembering and forgetting.

The Role of How Aeronautical Knowledge Is Initially Learned

It is widely known that memory for any type of learned material is affected by how that material is initially studied and learned. Memory for material memorized and rehearsed in a rote manner is weaker than for material that is processed in more elaborate and meaningful ways (Craik & Lockhart, 1972; Frase, 1975; Stein & Bransford, 1979). There is already considerable evidence that pilot applicants sometimes engage in crude memorization exercises to expedite passing of FAA pilot knowledge tests, at the expense of true understanding of the material (Casner, Jones, Puentes, & Irani, 2003). This suggests the possibility of a more dire interpretation of the test scores we observed. It may be that some pilots never truly understood the material probed by our questions, and therefore had difficulty recalling it as a collection of rote-memorized facts.

This hypothesis raised the question of how best to teach aeronautical knowledge. A common criticism of current aviation teaching methods is that many areas of aeronautical knowledge are taught in the abstract, that is to say, outside of the context in which they are to be used. Throughout the history of education, there have been repeated attempts to implement the practice of teaching knowledge and skills in specific and practical contexts (Dewey, 1938; Lave, 1988). In each case, a perhaps overzealous attempt to implement context-specific learning has arguably led to a later deterioration in fundamental skills. In each case, the teaching-in-context movement was later overthrown by a more conservative "back-to-basics" movement (Ravitch & Finn, 1989; Bloom, 1988). As Carroll (1990) pointed out, education is "frequently subject to headline whims and demagoguery (p. 1)" and "is chronically subject to trends (p. 1)." It is not clear whether or not recent attempts to reincarnate this idea, under the name of 'scenario-based training,' will avoid these same pitfalls.

References

Anderson, J. R. (2000). *Cognitive psychology and its implications*. New York: Worth Publishers.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Baddeley, A. (1998). *Human memory: Theory and practice*. Boston: Allyn and Bacon.

Blessing, S. B. & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 576-598.

Bloom, A. (1988). *The closing of the American mind*. New York: Simon and Schuster.

Carroll, J. M. (1990). *The Nurnberg funnel: Designing minimalist instruction for practical computer skill*. Cambridge, MA: The MIT Press.

Casner, S. M., Jones, K. M., Puentes, A., & Irani, H. (2003). FAA pilot knowledge tests: Learning or rote memorization? *International Journal of Applied Aviation Studies 3*(2), 277-289.

Casner, S. M. & Larkin, J. H. (1989). Cognitive efficiency considerations for good graphic design. *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, Ann Arbor, Michigan.

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

Dewey, J. (1938). *Experience and education*. New York: Collier Books.

Federal Aviation Administration (2006). *Federal aviation regulations*. Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration (2003). Personal communication.

Federal Aviation Administration (1999). *Aviation instructor's handbook. FAA-H-8039-9*. Washington, D.C.: Federal Aviation Administration.

Federal Aviation Administration (1991). Currency and additional qualification requirements for certificated pilots. *Advisory Circular No. 61-98A*. Washington, D.C.: Federal Aviation Administration.

Flouris, T. (2001, October). *The impact of ground schools in a collegiate aviation program on FAA written exam scores.* Paper presented at the University Aviation Association Annual Conference, Nashville, TN,.

Frase, L. T. (1975). Prose processing. In G. H. Bower (Ed.), *The psychology of learning and motivation Vol. 9*. (pp. 1-47). New York: Academic Press.

Gardner, B. (2002). *The complete private pilot*. Newcastle, WA: Aviation Supplies & Academics.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology 6*(2), 270-292.

Jeppesen Sanderson Inc. (1999). *Private pilot manual*. Englewood, CO: Jeppesen Sanderson Inc.

Kershner, W. K. (2001). *The student pilot's flight manual*. Ames, IA: Iowa State Press.

Koedinger, K. R. & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science 14*, 511-550

Lave, J. (1988). *Cognition in practice*. Cambridge: Cambridge University Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review 95*(4), 492-527.

Newell, A. & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.

Ravitch, D. & Finn, C. E., Jr. (1989). *What do our 17-year-olds know: A report on the first national assessment of history and literature*. New York: Harper-Collins.
Stein, B. S. & Bransford, J. D. (1979). Constraints on effective elaboration: Effects of precision and subject generation. *Journal of Verbal Language and Verbal Behavior 18*, 769-777.
Yates, F. A. (1966). *The art of memory*. The University of Chicago Press.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Gender Issues on the Flight-Deck: An exploratory analysis

Alexandra Kristovics
and
Jim Mitchell
University of Western Sydney
35 Stirling Highway
Crawley WA
Australia

Leo Vermeulen
University of Pretoria
Hillcrest, Pretoria
South Africa

Jan Wilson
Independence Air
Washington DC
USA

and

Monica Martinussen
University of Tromsø
Tromsø, Norway

Abstract

Aviation is a historically masculine occupation and the under-representation of female pilots in aviation led to the present research on possible gender bias among male pilots. A newly developed 4-factor measure of gender attitude was examined for construct and differential validity. Using data from 2009 pilots from the United States of America, South Africa, Australia and Norway, the 4-factor AGAQ was found to have adequate goodness of fit indices when using confirmatory factor analysis. Females scored higher than males on all factors, and instructors scored higher than pilots did on three factors. There were also differences in scores among pilots from different countries. Results also indicated that the opportunity to fly with the opposite gender and completing Crew Resource Management training also tended to increase scores leading to a more positive perception of female pilots. The benefits, and the limitations, of the measure are discussed.

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.

Introduction

Aviation is the ultimate global industry. Professional pilots tend to follow a career path that leads many of them into employment that crosses national and international boundaries. They carry with them their personal and professional perceptions that influence their behaviours (Rollinson, Broadfield, & Edwards, 1998). Within the aviation industry, however, the professional pilot carries with them perceptions that are historically masculine. These perceptions may affect relationships within the flight deck, particularly when the other pilot is female (Bateman, 1987).

There is little argument that female pilots are underrepresented in the aviation industry. For example, Helmreich and Merritt (1998, p. 42) estimated that, in 1996, only 3% of airline pilots in the United States were female. In Australia today, the majority of licences are held by male pilots. Information provided by the Civil Aviation Safety Authority to the Australian Transport Safety Bureau (Australian Transport Safety Bureau, 2005) showed that in 2004 only 857 (5.48%) of the 15,643 people holding a Private Pilot Licence were females. From the 5,088 of those holding a Commercial Pilot Licence, only 276 (5.42%) were females, and from the 6,552 holding an Air Transport Crew Licence, only 182 (2.78%) were females. However, some attempts have been made to address this imbalance. For example, Karp et al. (2002) have examined training methods to accommodate different male and female learning styles in an attempt to retain female pilots within the industry. The concern of retaining female pilots results from a need for more pilots to meet the demands of a growing market and to replace retired pilots (Karp et al., 2002).

As with many other industries, the aviation industry is bound by Affirmative Action (AA), Equal Opportunity (EEO) legislation and associated organisational policies. The effectiveness of these laws and policies, however, has been hotly debated. There is some evidence of adverse consequences in that women, who felt that they were beneficiaries of the policy, had doubts about their competence and lacked motivation to progress. There was also the stigma of incompetence and 'high visibility' attached to those who were viewed as being hired because of affirmative action policies (Muchinsky, 2003).

Whether females are beneficiaries of these policies or not, there may still be difficulties associated with female pilots' working relationship with male pilots. For example, Davey and Davidson (2000) revealed that isolation, sexism, and harassment are still being experienced by female pilots. Some female pilots have been shown to 'adapt' to this masculine culture by laughing at sexist jokes, going for drinks with the 'boys', and feeling as if they have to perform far beyond that of males (Davey & Davidson, 2000). These relationships may pose problems in maintaining female pilots within the aviation industry.

Little research has been undertaken on male pilot attitudes towards female pilots. Davey (2004), however, examined the discourse between instructors and trainee female pilots at an international airline and flying college in Europe. In general, the comments related to female trainee pilots as more responsible and less inclined to take risks than male pilots. Other comments related to females being more sympathetic, someone with whom others can talk to about their problems. Davey (2004) does admit that these comments reflected stereotypical ideas about females. Others, however, might classify these comments as reflecting 'benevolent sexism', "a subjectively favourable, chivalrous ideology that offers protection and affection to women who embrace conventional roles" (Glick & Fiske, 2001, p. 109). Indeed, Davey (2004, p. 644) stated that the instructors viewed those female cadet pilots more positively when they "conform to feminine ideals in terms of their body image, sexuality, and behaviour (for example, by acting as confidants)". Although there was an element of stereotypical ideals, Davey (2004) agreed that instructors did value the female cadet pilots because they possessed skills (communication, reliability etc) that were in line with the changing culture of a 'macho', risk-taking one to one where human factors such as teamwork, communication, rational decision-making and keeping calm in emergencies are becoming more desirable.

Although aviation is the ultimate global industry, there is some indication that males across different cultures hold a stereotypical bias of women in male-dominated positions. Using the example of management, Schein, Mueller, Lituchy and Lui (1996) quote Antal and Izraeli (1993, p. 63) who stated, "probably the single most important hurdle for women in management in all industrialized countries is the persistent stereotype that associates management being male." Schein et al. (1996) further explored this concept by examining how sex role stereotypes related to characteristics that were perceived as male or female using Chinese and Japanese samples. They then compared these results to those they obtained in 1992 (Schein & Mueller, 1992 in Schein et al., 1996). They found that males across all the countries that were examined, in China, Japan, Germany, the United Kingdom and the United States of America, perceived middle managers as "possessing the characteristics, attitudes and temperament more commonly ascribed to men in general than women in general" (1996, p. 38). There were, however, some differences amongst the females. The results of females from the United States of America indicated that they perceived males and females as equally likely to have the necessary management characteristics. The other nationalities had varying degrees of resemblance between descriptions of women in general and descriptions of managers. Although these studies were undertaken in 1992 and 1996, they used samples of business and management students who would then enter the workforce. One then, may question whether these findings, especially those of the male samples, would continue to perpetuate to the present decade. Management, however, does not have the same developmental history as that of aviation, or more specifically, that of an airline pilot. Along with architecture, airline pilot was the occupation that was strongly sex-stereotyped as males, and secretary and hairdresser for females (Miller & Budd, 1999 in Miller & Hayward, 2006). The perception of airline pilot as a 'male' occupation may arise from initial recruits being taken from the military and, therefore, sharing some of its culture of toughness, perseverance, risk-taking and living for today (Davey, 2004). Therefore, the bias against female pilots may suffer the same bias as that shown by males for female managers.

An examination of attitudes of males towards female pilots may also be relevant to manage more effectively communication and teamwork, that is, part of the human factor aspect of safety management. The present research plans to address this issue. Attitudes amongst pilots, and its implication to safety behaviours, have already been examined by Helmreich and his associates (eg. Merritt & Helmreich, 1996; Helmreich & Merritt, 2000). They explored the national, organisational, and professional culture and how these may relate to safety behaviours (Helmreich & Merritt, 2000). Attitudes, based on gender, may be either ameliorated or enhanced, depending on the norms displayed by national and organisational cultures, let alone the professional culture.

The present research has two aims. The first is to examine the construct validity and differential validity of a newly-developed 4-factor scale measuring perceptions of female pilots. The 4-factor structure was developed from the original 72-item Aviation Gender Attitude Questionnaire (AGAQ) developed by Vermeulen, Wilson and Mitchell (2004). Although the AGAQ represents a measure of gender attitude that is specific to pilots, there are some limitations to the measure. One of the limitations of the measure lies in the conceptual aspects related to the wording and coding of the items. Some of the items used in the original measure and, thus, incorporated into the 4-factor structure were reverse coded in an attempt to produce a measure where high scores indicated a more positive attitude and low scores indicated a more negative attitude towards female pilots. However, in doing so, there may be problems in meaning. For example, when the item 'Male pilots tend to 'take charge' in flying situations more than female pilots' is reverse coded, it does not necessarily imply that the opposite in meaning is achieved. If the original response was 'strongly disagree' to that item, the response might imply that females 'take charge' more than males (thus implying the opposite) or that males and females have a similar propensity to 'taking charge'. Unfortunately, this aspect of reverse coding may negate some of the 'real' attitudinal data, thus, masking some of the more important aspects of this attitude survey, and the consequent research results. Notwithstanding this limitation, the results do offer an indication of the perceptions that male pilots have towards female pilots. This research would then provide a basis for further exploration on this issue.

The second aim is to examine the relationship between these four factors and other variables, such as gender, nationality, certification (e.g., pilot or instructor), crew management training, and the opportunity pilots had to fly with the opposite gender. The paucity of the literature and research in this area is overdue, and the results of the present research may have practical implication for organisations in recruitment, training, and diversity management.

Method

Participants

The participants consisted of 2009 pilots whose nationalities were Australians (53%), South Africans (28%), Americans (9%), and Norwegian (10%). The data was collected from various flight organisations in Australia, South Africa, Norway, and United States of America. Military pilots, except for the Norwegian cohort, were not surveyed, as access was not granted. From this sample, 1064 participants were Australians whose mean age was 44.73 (SD=11.98) and where 937 (87.5%) were males. The mean age of the 562 participants whose nationality was South African was 37.08 (SD=11.28) years, and consisted of 534 (94.0%) males. The 181 participants from the United States of America had a mean age of 46.21

(SD=12.70) and 52 (27.1%) were males. The Norwegian pilots (202) mean age was 36.82 (10.1%). The mean number of years of flying experience for all participants was 18.38 (11.51) years and the mean flying time was 675.56 (6012.21) hours. While overall 2122 pilots participated in the study, those with different nationalities from the ones mentioned above (113) were removed from each nation's data. The data from participants in South Africa was initially utilised by Vermeulen, Wilson, and Mitchell (2004) for their 43-item 2-factor model. The present study, however, examines the additional newly gathered data from Australia, Norway, and the United States of America.

Measure

The 15-item 4-factor scale was developed from the 43-item measure of gender attitude (Aviation Gender Attitude Questionnaire – AGAQ) developed by Vermeulen, Wilson and Mitchell (2004). Initially, the AGAQ consisted of 72 items, but a series of exploratory factor analyses revealed an initial analysis of 14 factors was not representative of the measure, that is, when examining the eigenvalues and the scree plot. The final solution then consisted of two factors comprising of 43 items, that is, Flying Proficiency and Safety Orientation. The scales showed good reliability of .96 for the former and .86 for the latter. The Aviation Gender Attitude Questionnaire (AGAQ) was designed as the basis for a cross-cultural study of gender-based issues in aviation and, specifically, attitudes towards female pilots. Respondents were required to rate each item on a Likert scale where 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree and 5 = strongly agree. Several items required reverse scoring, and where high scores indicated a positive attitude and low scores indicated a negative attitude towards female pilots. The survey also consisted of demographic data such as gender, nationality, aircraft certification (e.g. private, commercial, and flight instructor), crew resource management training, and the opportunity to fly with the opposite gender.

Rather than the determination of factors being statistically driven, one may also observe the conceptual issues that may be inherent in items from questionnaires. These concepts can then be examined statistically for goodness of fit, that is, how well the proposed model fits the data. When examining the original 72 items of the AGAQ at face value, a number of specific facets of gender attitudes could be determined. Some of these included leadership and decision-making, emotionality and assertiveness, issues related to affirmative action and dangerous flying behaviour. These four facets were named Decision/Leadership, Assertiveness, Hazardous Behaviour, and Affirmative Action. The first three factors reflect the dimensions that are considered valuable qualities in pilots, that is, the human factors in piloting skills (Davey, 2004). The Affirmative Action factor relates to attitudes of female pilots who may be perceived as being beneficiaries of policies associated with Affirmative Action and Equal Opportunity. It is also related to the possibility of perceptions that flight programmes have been relaxed in response to fulfilling quotas of females into the industry.

The 4-item Decision/Leadership factor consisted of items such as 'female pilots often have difficulty making decisions in urgent situations'; the 4-item Assertiveness factor consisted of items such as 'male pilots tend to take charge in flying situations more than female pilots' (reverse scored); the 4-item Hazardous Behaviour factor consisted of items such as ‘male pilots are more likely to run out of fuel than female pilots’; and the 3-item Affirmative Action factor consisted of items such as ‘flight training standards have been relaxed so that it is easier for women to get their wings’. Table 1, below, lists these items.

Procedure

The data was gathered from male and female pilots holding current and valid aerial licences in their respective countries. In the United States, the AGAQ was made available on the website for data collection (www.aviatrices.org), the International Society of Women Airline Pilots (www.iswap.org), the quarterly magazine of The Ninety-Nines (Waypoint) and distributed electronically and in printed format to various military, professional, and private pilots. In South Africa, the questionnaires were distributed to various airlines, pilot training academies, and charter companies. Pilots in Australia were contacted through their various pilot associations. These are the Aircraft Owners and Pilots Association (AOPA), Australian Federation of Air Pilots (AFAP), Australian and International Pilots Association (AIPA,) and the Australian Women Pilots Association (AWPA). Norwegian respondents were contacted through the military, airlines, and a pilot training company.

Statistical Analyses

Unlike exploratory factor analysis, one of the advantages of structural equation modelling (SEM) is that it enables the researcher to postulate relations between the observed measures and the latent variables a priori. This a priori relationship between the observed variables and the latent variable would then be evaluated statistically to determine its goodness of fit to the data (Jöreskog, 1993). Rather than using statistical analysis to derive more homogenous factors, a number of items were chosen at face value to represent different facets of the measure. LISREL 8.5 was used to determine the Goodness of Fit Indices for the model and factors, and SPSS 11.5 was used to evaluate the different relationships between these factors and demographic variables.

Results

The first analysis undertaken was to estimate the Goodness of Fit Indices (GFIs) for the 4-factor model. GFIs indicate how well the proposed model fits the data. GFIs where the GFI, NNFI, CFI are close to 1.00 and the RMSEA that is closer to .05 are indicative of a good fit, although .08 for the latter would still represent a reasonable fit (Byrne, 1998). Table 1 lists these items.

Table 1
Items from the 4-factor solution

Item	Factors
	Decision-Leadership
11.	Female pilots often have difficulty making decisions in urgent situations * (1)
19.	The likely reason for accidents involving women pilots is poor decision- making * (1)
51.	Male pilots tend to be more rational in making decisions than female pilots *

55.	Female pilots' decision-making ability is as good in an emergency situation as it is in routine flights. (1)
	Assertiveness
15.	Male pilots tend to 'take charge' in flying situations more than female pilots * (3)
24.	Male pilots are less nervous when piloting than female pilots * (1)
43.	Male pilots tend to be more *confident* than female pilots * (1)
49.	Male flight students tend to be less fearful of learning stall procedures than female students (1)
	Hazardous Behaviour
58	Male pilots are more likely to run out of fuel than female pilots (2)
62	Male pilots are more likely to land with the landing gear up than female pilots (2)
67	Male pilots tend to take greater risks than female pilots (2)
70	Female pilots tend to practice more situational awareness than male pilots (2)
	Affirmative Action
40	Professional female pilots are only in positions they are in because airlines want to fulfil affirmative action quotas * (3)
52	Flight program standards for the airlines/military have been relaxed in order to increase the number of female pilots * (1)
69	Flight training standards have been relaxed so that it is easier for women to get their 'wings' * (1)

NB: Item numbers relate to those items taken from the original 72-item AGAQ. Asterisks indicate that the item was reverse scored; (1) indicates that the item was included in the original two-factor model on Flying Proficiency; (2) indicates that the item was included in the original two-factor model on Flying Proficiency; (3) indicates that the item was not included in the original two-factor model.

Analyses were undertaken for those groups that had sufficient sample sizes for the analyses, that is, with data from those participants whose nationalities were Australian and South African. Analyses were also undertaken for the whole sample. However, one of the factors to be examined, Affirmative Action, consisted of only three items. Unfortunately, for a measure that consists of only three items, there will be a one-to-one correspondence between the data and the structure parameters, resulting in a just-identified model that leaves no degrees of freedom. Therefore, goodness-of-fit indices cannot be undertaken on a just-identified model. As can be seen from Table 2, the GFIs for the congeneric and 4-factor model indicated that the model had a good fit to the data for Australians, South Africans and the whole sample, including the participant from Norway and the United States of America.

Table 2
Goodness of Fit Indices for the 4-factor and 2-factor models of the AGAQ

	N	χ^2	df	GFI	NNFI	CFI	RMSEA
Australia	1064						
Decision/Leadership		2.02	2	1.00	1.00	1.00	.02
Assertiveness		3.04	2	1.00	1.00	1.00	.02
Hazardous Behaviour		2.86	2	1.00	1.00	1.00	.01
4-Factor Solution		437.52	84	.95	.95	.96	.06
South Africa	562						
Decision/Leadership		2.80	2	1.00	1.00	1.00	.03
Assertiveness		.60	2	1.00	1.00	1.00	.01
Hazardous Behaviour		4.24	2	1.00	1.00	.99	.04
4-Factor Solution		350.22	84	.92	.92	.94	.07
South Africa	2009						
Decision/Leadership		4.70	2	1.00	1.00	1.00	.03
Assertiveness		7.06	2	1.00	1.00	1.00	.03
Hazardous Behaviour		13.59	2	1.00	.99	1.00	.05
4-Factor Solution		628.78	84	.96	.96	.97	.06

Figure 1 displays the parameters for the total sample. As can be seen from Figure 1, each item adequately represents the underlying construct, as seen by the errors on the left of the figure, and all parameter estimates were significant. The correlations of the latent factors can be seen in Table 3.

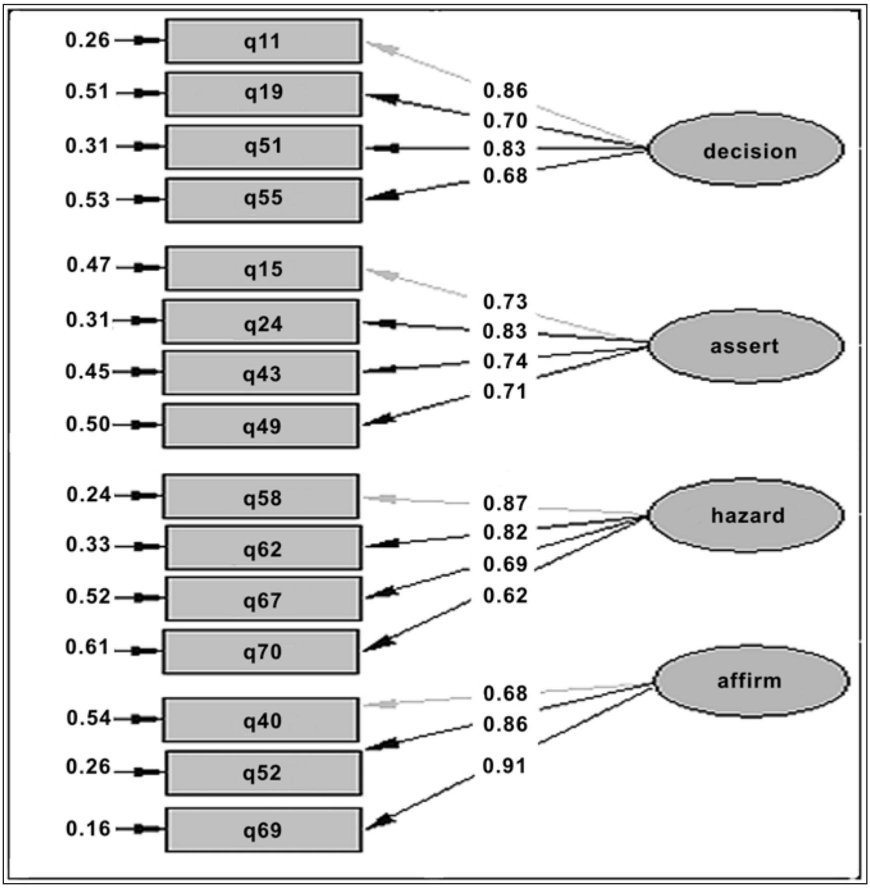


Figure 1. Parameter estimates of the 4-factor model (Decision/Leadership, Assertiveness, Affirmative Action, and Hazardous Behaviour) for the total sample.

Table 3
Correlation between the latent factors from the 4-factor model.

	1	2	3	4
1. Decision/Leadership	1.00			
2. Assertiveness	.86	1.00		
3. Hazardous Behaviour	- .16	- .38	1.00	
4. Affirmative Action	.76	.65	-.08	1.00

The Cronbach reliability coefficients for Decision/Leadership, Assertiveness, Hazardous Behaviour and Affirmative Action are .83, .76, .78 and .81 respectively for the Australian sample; .78, .80, .72 and .79 respectively for the South African sample; .71, .75, .84 and .72 respectively for the American sample; .75, .80, .81 and .74 respectively for the Norwegian sample; and .81, .79, .79 and .81 respectively for the total sample.

The relationship between the factors from the 4-factor model and the 2-factor model was also examined by summating the scores of the items that represented each of these variables. Using Pearson's alpha, results showed Decision/Leadership, Assertiveness and Affirmative Action have strong correlations with Flying Proficiency from the 2-factor model and Affirmative Action have strong correlations with Flying Proficiency (.89, .78 and .73 respectively) from the 2-factor model, and Hazardous Behaviour only has a strong correlation with Safety Orientation (.86) from the 2-factor model (Table 4).

Table 4
Pearson correlations between the variables from the 2- and 4-factor models.

	1	*2*	*3*	*4*	*5*	*6*
1. Decision/Leadership	*1.00*					
2. Assertiveness	*.67*	*1.00*				
3. Hazardous Behaviour	*-.09*	*-.32*	*1.00*			
4. Affirmative Action	*.61*	*.52*	*-.04*	*1.00*		
5. Flying Proficiency	*.89*	*.80*	*-.18*	*.74*	*1.00*	
6. Safety Orientation	*-.14*	*-.42*	*.86*	*.07*	*-.23*	*1.00*

To determine the relationship between the factors from the 4-factor solution with other variables, such as gender, nationality, certification, Crew Resource management Training and flying hours with the opposite gender, a series of t-tests and ANOVA were performed. Most of these analyses were also undertaken with original 2-factor solution to identify whether any differences emerged from those relationships found with the 4-factor solution, that is, to explore whether the 4-factor solution had better differential validity than the 2-factor solution. The assumptions of ANOVA and t-tests were deemed satisfactory. The means of the scores on the six factors for males and females from each country can be seen in Table 5.

Table 5
Means of males and females for variables from the 4- and 2-factor models

	N	*Decision/ Leadership*	*Assertive-ness*	*Hazardous Behaviour*	*Affirmative Action*	*Flying Proficiency*	*Safety Orientation*
Australia							
Males	*933*	*3.42 (.74)*^{b}	*3.04 (.75)*^{b}	*2.61 (.68)*^{a}	*3.52 (.86)*^{b}	*3 45 (.61)*^{b}	*2.87(.59)*^{ab}
Females	*131*	*4.27 (.57)*$_{a}$	*3.35 (.73)*$_{b}$	*3.03 (.76)*$_{a}$	*4.25 (.58)*$_{ab}$	*4.16 (.43)*$_{a}$	*3.32 (.63)*$_{b}$
South Africa							
Males	*531*	*3.16 (.75)*^{c}	*2.69 (.77)*^{c}	*2.66 (.70)*^{a}	*3.15 (.92)*^{c}	*3.23 (.61)*^{c}	*3.06 (.59)*^{a}
Females	*31*	*4.24 (.55)*$_{a}$	*3.30 (.84)*$_{b}$	*3.25 (.77)*$_{a}$	*4.19 (.63)*$_{b}$	*4.16 (.44)*$_{a}$	*3.67(.55)*$_{b}$

American							
Males	*42*	*3.49 (.83)*^{b}	*3.10 (.78)*^{b}	*2.54 (.93)*^{a}	*3.20 (1.18)*^{c}	*3.57 (.69)*^{b}	*2.74 (.85)*^{bc}
Females	*139*	*4.21 (.66)*$_{a}$	*3.41 (.87)*$_{b}$	*3.35 (.79)*$_{a}$	*4.20 (.67)*$_{b}$	*4.12 (.54)*$_{a}$	*3.47(.67)*$_{b}$
Norway							
Males	*191*	*3.85 (.74)*^{a}	*3.52 (.86)*^{a}	*2.23 (.84)*^{b}	*4.26 (.67)*^{a}	*3.98 (.61)*^{a}	*2.55 (.71)*^{c}
Females	*11*	*4.11 (.72)*$_{a}$	*4.02 (.69)*$_{a}$	*2.02 (.75)*$_{b}$	*4.64 (.41)*$_{a}$	*4.35 (.54)*$_{a}$	*2.41 (.77)*$_{a}$
Total							
Males	*1697*	*3.39 (.83)*	*2.99 (.81)*	*2.58 (.72)*	*3.48 (.93)*	*3.44 (.65)*	*2.89(.63)*
Females	*312*	*4.24 (.61)*	*3.39 (.81)*	*3.16 (.81)*	*4.24 (.63)*	*4.15 (.49)*	*3.39 (.68)*

NB. Different superscripts represent significant differences between males from the different countries for each of the factors.
Different subscripts represent significant differences between females from the different countries for each of the factors.

Gender

Results from the t-test for the combined group showed that there was a significant difference between the genders for all the variables from both the 2-factor and the 4-factor solutions. Specifically, the results for Decision/Leadership were t(2007)=-21.35, p<.001; t(2120)=-8.17, p<.001; for Assertiveness; t(2007)=-12.73, p<.001; for Hazardous Behaviour; and t(2007)=-13.83, p<.001for Affirmative Action. Results for Flying Proficiency and Safety Orientation were, t(2007)=-22.34, p<.001 and t(2007)=-12.70, p<.001 respectively. Females scored higher on all factors than males.

Gender by Nationality

To determine which gender differences were due to different nationalities, a series of 2 (gender) X 3 (nationality) between-subjects ANOVAs were performed. Unfortunately, there were only 11 females in the Norway sample whose results may not be indicative of that population. Therefore, the following analyses were undertaken for the American, Australian, and South African samples only.

For the Decision/Leadership analysis, results indicated that only the main effect of gender was significant, gender, F(1,1801)=176.52, p<.001, partial η^2=.09. The main effect of nationality and the gender by nationality interaction were not significant. Females scored higher than males, and there was no significant difference between the different nationalities. For the Assertiveness analysis, results indicated that both the main effects were significant, but not the interaction: gender, F(1,1801)=34.67, p<.001, partial η^2=.02: nationality, F(2,1801)=4.17, p<.05, partial η^2=.01. Female scores were similar; all of them scoring higher than males, but the South African males scored considerably lower than the Americans and Australians males.

For the Hazardous Bahaviour analysis, results indicated that only the main effect of gender and the interaction were significant gender, F(1,1801)=89.59, p<.001, partial η^2=.05: gender by nationality, F(2,1801)=4.05, p<.05, partial η^2=.01. The scores of American males and females differed significantly from the

South African and Australian male and female scores. American males scored lower than the males from the other two countries, and American females scored higher than the females from the other two countries. The interaction effects can be seen in Figure 2.

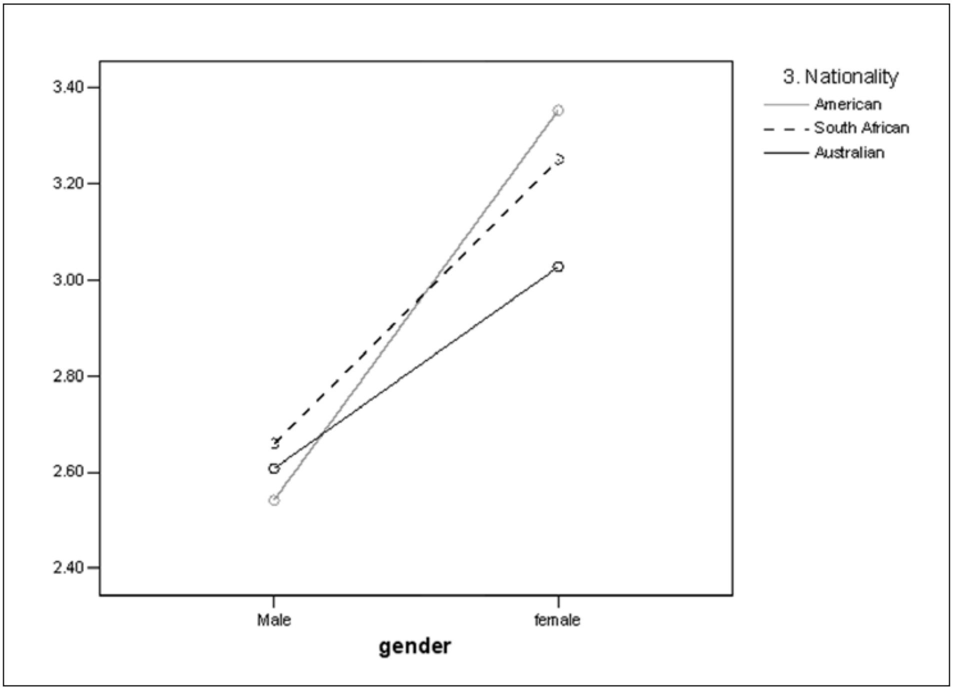


Figure 2. Gender and Nationality interaction effect for Hazardous Behaviour.

For the Affirmative Action analysis, results indicated that both the main effects were significant, but not the interaction: gender, $F(1,1801)=141.93$, $p<.001$, partial $\eta^2=.07$ and nationality, $F(2,1801)=4.51$, $p<.05$, partial $\eta^2=.01$. All females scored higher than all males, and Australians scored higher than those from the other two countries.

For the Flying Proficiency analysis, results indicated that only the main effect of gender and the interaction were significant: gender, $F(1,1801)=186.72$, $p<.001$, partial $\eta^2=.09$: gender by nationality, $F(2,1801)=2.14$, $p<.05$, partial $\eta^2=.01$.

Scores of females were higher than that of the males, and South African males scored lower than the males from the other two countries. The females from South Africa, however, scored the same as the females from the other two countries. The interaction effects can be seen in Figure 3.

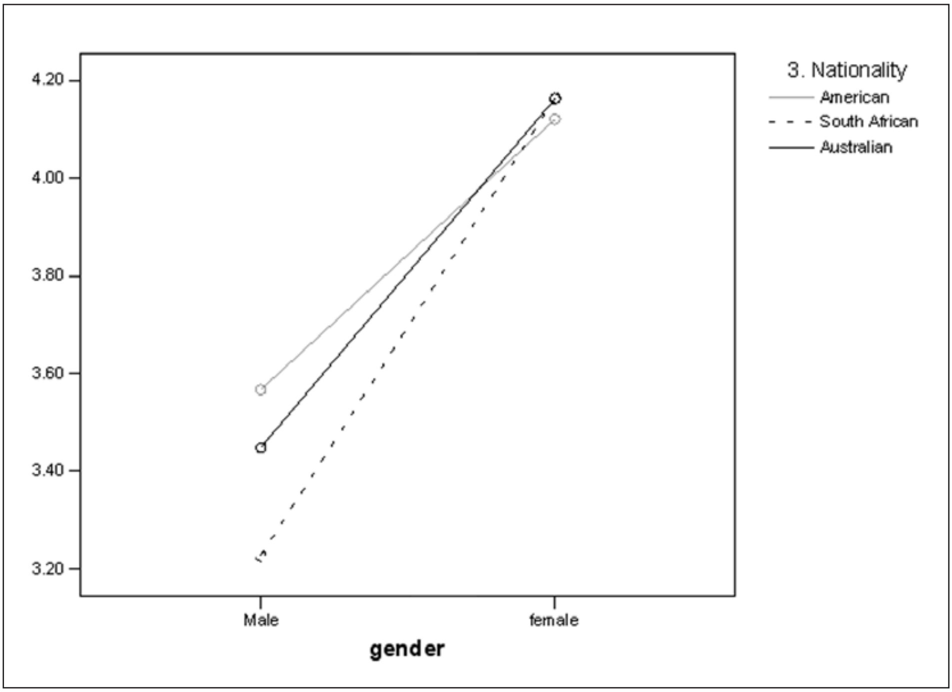


Figure 3. Gender and nationality interaction effect for flying

For the Safety Orientation analysis, results indicated that both the main effects were significant, but not the interaction: gender, $F(1,1801)=117.67$, $p<.001$, partial $\eta^2=.06$: nationality, $F(2,1801)=3.60$, $p<.001$, partial $\eta^2=.01$. However, the interaction for this variable achieved near significance, where $p=.055$. Females scored higher than males, and both males and females from South Africa scored higher than males and females from Australia and the United States of America. The males from the United States of America scored lower than those from Australia, but the American females scored higher than the Australian females. The interaction effects can be seen in Figure 4.

In summary, the main effect of gender was shown to be significant for each of the factors examined, but where the main effect of nationality was seen for Assertiveness, Affirmative Action, and Safety Orientation. The interaction, however, was significant for the variables of Hazardous Behaviour, Flying Proficiency, and Safety Orientation.

Opportunity to fly with the opposite gender

The opportunity to fly with the opposite gender was also examined, that is, to establish whether professional exposure reduced negative perceptions of male pilots to female pilots. Because there were relatively fewer female pilots than male pilots, males had disproportionate fewer opportunities to fly with female pilots. For those from all nationalities who responded to this item, 190 males stated never had the opportunity to fly with the opposite gender, compared 1332

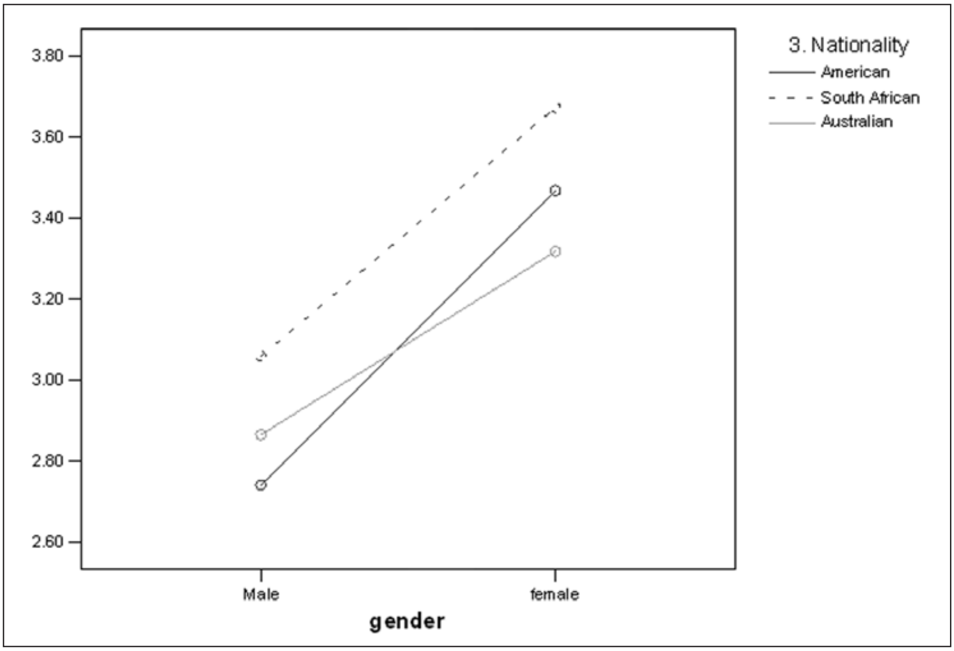


Figure 4. Gender and Nationality interaction effect for Safety Orientation.

who stated rarely, 136 who stated sometimes, 28 who stated often and 6 who stated mostly. Females displayed the opposite pattern. For those from all nationalities who responded to this item, six females stated they never had the opportunity to fly with the opposite gender, compared to 32 who stated rarely, 32 who stated sometimes, 76 who stated often, and 166 who stated mostly. Therefore, to be able to obtain enough number of participants for each cell of the analysis, only those from the South African and Australian samples were included. Further, each gender was divided into two groups. Group 1 consisted of those who never or rarely flew with the opposite gender and Group 2 consisted of those who sometimes, often, and mostly flew with the opposite gender.

Results from the 2 (opportunity to fly with others) X 2 (South Africa and Australia) analysis of variance for the Decision/Leadership indicated that both nationality and flying with the opposite gender were significant: $F(1, 1618)=58.67$, $p<.001$, partial $\eta^2=.04$ for nationality; $F(1, 1618)=35.81$, $p<.001$, partial $\eta^2=.02$ for flying opportunity. Results were also statistically significant for the interaction, $F(1, 1618)=6.17$, $p=.01$, partial $\eta^2=.01$. The means for South Africa and Australia can be seen in Table 6. Australia scored more than South Africa, and the scores for Australia increased more than that of South Africa, the more they had the opportunity to fly with female pilots.

Results from Assertiveness analysis indicated that only nationality was significant, $F(1, 1618)=58.67$, $p<.001$, partial $\eta^2=.04$. Results were not statistically significant for flying with females, ($F(1, 1618)=2.83$, $p>.05$) nor the interaction ($F(1, 1618)=3.28$, $p>.05$). Australia scored higher than South Africa both where there was or was not an opportunity to fly with the opposite gender, but the opportunity to fly with females did not result in a significant change of scores for both nationalities.

Results from the Hazardous Behaviour analysis indicated that flying with the opposite gender was significant: $F(1, 1618)=45.61$, $p<.001$, partial $\eta^2=.03$. Results were not statistically significant for nationality, ($F(1, 1618)=.01$, $p>.05$) nor the interaction ($F(1, 1618)=1.15$, $p>.05$). Both Australia and South Africa showed an increase in scores the more they flew with female pilots. There was no significant difference between the countries on their scores.

Results from the Affirmative Action analysis indicated that both flying with the opposite gender and nationality were significant: $F(1, 1618)=22.15$, $p<.001$, partial $\eta^2=.01$ and $F(1, 1618)=61.44$, $p<.001$, partial $\eta^2=.04$ respectively. Results were not statistically significant for the interaction ($F(1, 1618)=2.03$, $p>.05$). These results indicated that Australia showed higher scores than South Africa, and each country showed similar increases in scores the more they flew with the opposite gender.

Results from the Flying Proficiency analysis indicated that both flying with the opposite gender and nationality were significant: $F(1, 1618)=33.86$, $p<.001$, partial $\eta^2=.02$ and $F(1, 1618)=68.59$, $p<.001$, partial $\eta^2=.04$ respectively. Results were also statistically significant for the interaction, $F(1, 1618)=9.97$, $p<.001$, partial $\eta^2=.01$. Similar to Decision/Leadership, Australian scored more than South Africa, and the scores for Australia increased more than that of South Africa, the more they had the opportunity to fly with female pilots.

Results from the Safety Orientation analysis indicated that both flying with the opposite gender and nationality were significant: $F(1, 1618)=74.18$, $p<.001$, partial $\eta^2=.04$ and $F(1, 1618)=10.53$, $p<.001$, partial $\eta^2=.01$ respectively. Results were not statistically significant for the interaction, $F(1, 1618)=2.04$, $p>.05$. These results indicated that South Africa showed higher scores than Australia, and each country showed similar increases in scores the more they flew with the opposite gender.

Table 6

Means of males for variables from the 4- and 2-factor models with opportunity to fly with the opposite gender

	N	Decision/ Leadership	Assertive- ness	Hazardous Behaviour	Affirmative Action	Flying Proficiency	Safety Orientation
Australia							
Never to rarely	878	3.45 (.75)	3.05 (.74)	2.59 (.67)	3.55 (.86)	3 47 (.60)	2.85 (.59)
Sometimes to Mostly	182	3.89 (.84)	3.23 (.77)	2.91 (.77)	3.92 (.85)	3.85 (.68)	3.25 (.63)
South Africa							
Never to rarely	453	3.19 (.73)	2.72 (.77)	2.64 (.69)	3.16 (.92)	3.25 (.61)	3.04 (.59)
Sometimes to Mostly	109	3.37 (.94)	2.71 (.84)	2.91 (.77)	3.36 (1.01)	3.37 (.77)	3.09 (.62)

In summary, the scores were higher for male pilots the more they had the opportunity to fly with the opposite gender. Australian males also scored higher than the South African males on Decision/Leadership, Assertiveness, Affirmative Action, and Flying Proficiency. South Africans, however, scored higher than Australians for Safety Orientation. There were no differences between the two nationalities for Hazardous Behaviour. Interaction effects were seen for Decision/ Leadership and Flying Proficiency only.

Certification – Pilots and Instructors

The main purpose of these analyses was to estimate whether pilots and instructors differed in their attitudes towards females in aviation using the variables from the 4-factor model. From the 1977 participants who responded to this question, the pilots in the analyses included Private Pilots, Commercial Pilots, Instrument-Rated Pilots, Multi-Engine Rated Pilots, and Airline Transport Pilots. Instructors consisted of Flight Instructors, Flight Instructors (Multi-Engine) and Flight Instructors (Instrument). Table 7 shows the number of males and females of Pilots and Instructors, and the means and standard deviations of each of the variables examined below. A series of t-tests were undertaken for each gender to examine whether male instructors perceived female pilots more positively than male pilots.

The following analyses relate to males only. For Decision/Leadership, there was a significant difference between the means where instructors scored higher than pilots, $t(1670)=-3.97$, $p<.001$. There was also a significant difference between the means for Assertiveness where instructors scored higher than pilots, $t(180.64)=-5.40$, $p<.001$ (correcting for inequality of variance). For Hazardous Behaviour, there was a significant difference between the means where pilots scored higher than instructors, $t(177.75)=2.06$, $p<.001$ (correcting for inequality of variance). For Affirmative Action, there was a significant difference between the means where instructors scored higher than pilots, $t(1670)=-3.52$, $p<.001$. For Flying Proficiency, there was a significant difference between the means where instructors scored higher than pilots, $t(181.20)=-5.89$, $p<.001$ (correcting for inequality of variance). For Hazardous Behaviour, there was a significant difference between the means where pilots scored higher than instructors, $t(176.89)=4.45$, $p<.001$ (correcting for inequality of variance).

Similar analyses were also undertaken for females. None of the results for the t-tests showed significant differences between pilots and instructors, except for Hazardous Behaviour. The results for Safety Orientation were $t(304)=2.12$, $p<.05$, where pilots scored higher than instructors.

In summary, male instructors scored higher than male pilots on all variables except Hazardous Behaviour and Safety Orientation where the opposite pattern was found, that is, instructors scored less than pilots. In contrast, female pilots and instructors showed no difference between the scores on all variables, except for Hazardous Behaviour. Similar to males, female instructors scored lower on this dimension than female pilots.

Table 7
Means of instructors and pilots for variables from the 4- and 2-factor models

	N	Decision/ Leadership	Assertive- ness	Hazardous Behaviour	Affirmative Action	Flying Proficiency	Safety Orientation
Males							
Pilots	1514	3.36 (.77)	2.95 (.78)	2.60 (.70)	3.44 (.92)	3.41 (.63)	2.91 (.61)
Instructors	158	3.62 (.84)	3.31 (.94)	2.45 (.89)	3.72 (1.00)	3.72 (.75)	2.68 (.79)
Females							
Pilots	270	4.25 (.62)	3.36 (.81)	3.20 (.80)	4.23 (.64)	4.14 (.49)	3.41 (.65)
Instructors	36	4.16 (.60)	3.56 (.87)	2.90 (.92)	4.24 (.53)	4.20 (.45)	3.27 (.83)

Crew Resource Management Training

Crew resource Management was examined by dividing the participants into those who had completed the training and those who did not. Analyses were undertaken separately on males and females. The results of the t-tests for males showed that there were no significant differences between those who had undertaken the training and those who had not for the following factors, Decision/Leadership, Affirmative Action and Flying Proficiency. There were significant differences between the groups for the following: Assertiveness, $t(1686)=2.40$, $p<.05$; Hazardous Behaviour, $t(1686)=-6.17$, $p<.001$; and Safety Behaviour, $t(1686)=-5.73$, $p<.001$. The means and standard deviations for these variables are as follows: Assertiveness, 3.01 (.78) with training and 2.88 (.80) with no training; Hazardous Behaviour, 2.53 (.70) with training and 2.81 (.76) with no training; and Safety Behaviour, 2.84 (.62) with training and 3.07 (.63) with no training. Analyses were also undertaken for females. The results showed that there were no significant differences between those who had undertaken the training and those who had not taken the training.

In summary, there was no difference between the scores between females who had or had not undertaken training on all factors from the 4- and 2-factor solutions. Males who had undertaken the training scored higher on Assertiveness than those males who did not have that training. The opposite pattern occurred for Hazardous Behaviour and Safety Orientation, that is, where those with training scored lower than those who did not have training.

Discussion

The GFIs for the 4-factor structure indicated that this model had a good fit to the data for Australians and South Africans and for the total sample. Further, the 4-factor model focuses on more specific facets of possible gender bias than the more generalised 2-factor structure. For example, nine of the 15 items of the new 4-factor structure were taken from the Flying Proficiency (from the 2-factor model). These nine items, however, were positioned on three factors, that is, 4 for Decision/Leadership, 3 for Assertiveness, and 2 for Affirmative Action. The correlations of these factors with Flying Proficiency indicate this relationship. The four items of Hazardous Behaviour were taken from the 12 items that made up Safety Orientation, and there was a strong correlation between the two. The results also indicate that the 4-factor structure is more parsimonious, and shows more differential validity than the 2-factor model. Parsimony relates to the 4-factor model consisting of 15 items compared to the 2-factor model consisting of 43 items.

Although these three factors from the 4-factor solution had high correlations with Flying Proficiency, they still showed differential validity when examining each of their relationship with gender and nationality. For example, the ANOVA analyses revealed that for Decision/Leadership, Assertiveness and Affirmative Action showed significant main effect of gender but Assertiveness showed a main effect of nationality. The analysis for Flying Proficiency, however, indicated that there was only a significant main effect of gender. The interaction effect was shown for Flying Proficiency, but this may be due to the influence of more items (31) being included, such as proneness to fatigue, sense of direction and other similar items. There were, however, too few of these items to make up separate, distinct factors. The 4-factor model, therefore, offers better differential validity and parsimony than the 2-factor model. As stated previously, differential validity was observed in the relationships between the different facets of the 4-factor model with gender and nationality interactions, as well as certification. In relation to the similarity of hazardous Behaviour and Safety Orientation, there were differences in the analyses for flying with the opposite gender and gender. The results for Hazardous Behaviour showed that flying with the opposite gender was significant, but not gender. In contrast, both gender and flying with the opposite gender were significant for Safety Orientation.

Although many of the analyses related to nationality were shown to be significant, the results should be accepted with caution. The strength of association for nationalities, as shown by partial η^2 in the results, was quite small. As stated by Tabachnick and Fidell (1996), tests of significance "do not test the degree to which the IV(s) and DV are related" (p. 53). They further state that the strength of association "assesses the proportion of variance in the DV that is predictable from the knowledge of the levels of the IV" (p. 53). They further add that some research produces results that are statistically significant but realistically meaningless *because* of the lack of strength of association. The analyses that related to gender, however, indicate that the strength of association may have implications for intervention to reduce possible gender bias.

Female pilots compared to males were more positive in their responses for all variables examined. There were also some differences on scores relating to males and females from the different nationalities, although those analyses related to nationality had a low strength of association (as stated above). The comparatively small number of people in the Norwegian and American samples may have contributed to this weakened result. The results, however, do point to some basis on which to develop culture-specific interventions to reduce gender bias. There have already been some findings that cultures based different degrees of 'power distance' and 'individualism-collectivism' having links to safety management. "National culture has been implicated as a contributory factor in analyses of air crashes" (Helmreich, 2000, p. 136), more specifically in how these variables may affect the way information is shared, including the willingness of subordinates to speak up with critical information. This situation may even be more problematic when the crew consists not only of pilots of different cultures, but different genders.

The opportunity to fly with the opposite gender' also seemed to have an impact in increasing scores for Decision/Leadership, Assertiveness, and Affirmative Behaviour from the 4-factor model. Social psychologists have long studied the concept of 'contact' as a way of reducing prejudice but the situation is more complex than just exposing two parties to each other (e.g., Allport, 1954; Miller & Brewer, 1984). Allport (1954) suggested four conditions on which contact could reduce prejudice. These are contact had to be sustained and involve close interaction; contact must produce cooperative interdependence; contact must promote equality (status); and the interaction is in an atmosphere favouring equality.

Undertaking Crew Resource Management training seems to have increased the scores of males on all factors of the 4-factor model, except for Hazardous Behaviour. This finding is difficult to interpret. The Crew Resource Management training is related to safety management and such training may bring to the fore the possible dangerous behaviour that has been performed, not just by females, but also by males. One of the limitations of the study is that there were no questions related to the perceptions of male pilots, thus enabling this suggestion to be examined. Another limitation is that there is no indication whether the training was undertaken in a mixed gender environment, where males may have formed attitudes that were more positive towards females. Unfortunately, the wording of these items poses difficulties. As stated above, when these items are reverse-scored, it does cause some confusion. For example, reverse-scoring the item 'Male pilots are more likely to run out of fuel than female pilots' does not imply that females are less likely to run out of fuel. It may also indicate that both males and females are unlikely (or likely) to run out of fuel.

The finding that instructors had higher scores than pilots on all factors of the 4-factor model, except for Hazardous Behaviour, is promising. The usual first point of contact for pilots is with instructors. Their more positive perceptions may enhance the learning situation for females. There also seemed to be a positive aspect in the relationship of Affirmative Action with factors such as Decision/Leadership and Assertiveness. The correlation for Affirmative Action for males only was .70 and .53 for Decision/Leadership and Assertiveness respectively. The positive correlations among these variables imply that policies of Affirmative Action encourage those females with the perceived necessary skills the opportunity to become pilots.

When one measures attitudes, however, one may need to incorporate other important factors. Attitudes are made up of three components, that is, emotional, behavioural and cognitive (Oskamp, 1977). The 72 items of the Aviation Gender Attitude Questionnaire (AGAQ), however, are mostly made up of the cognitive/belief component. More items that may be indicative of the affective and behavioural component may be more predictive in estimating the relationship between gender bias and other variables, thus enabling the aviation industry to put forward more effective cost-benefit programmes. Indeed, research by Davey and Davidson (2000) showed that it was more of the behavioural component that was apparent in sexist attitudes within the aviation industry. These consisted of behaviours such as playing practical jokes, pranks or teasing. Furthermore, the classic work by LaPierre (1934, in Dockery & Bedeian, 1989) found that cognitions did not always relate to behaviour. More recently, Cacioppo, Gardner, and Berntson (1997) proposed that attitudes should emphasizes their "motivational substrates and, in particular, the distinction between the motivational processes underlying positive and negative evaluation processes" (p. 5). In other words, positive and negative evaluations can bring forward more definitive results of attitudes, that is, positive or negative attitudes towards an object, as well as ambivalence or neutrality/indifference.

Proper measurement is an essential component for any benchmarking strategies. As Becker, Huselid and Ulrich (2001) pointed out, sound measurement does two things. Firstly, valid measures help give focus on aspects of the organisation that create value. Secondly, "it provides a valid and systematic justification for resource allocation decisions" (p. 111). With more organisations paying increased attention to human factors (especially in safety behaviours), proper measures and research is essential. The AGAQ is a good beginning in measuring aspects of gender attitudes in the aviation industry. The poor relationships with other variables and the problems associated with the conceptual aspect of reverse-coding, however, point towards the need to further improvement of this measure. Once this has been undertaken, it may then be a measure that can be more efficiently used as a basis for more cost-benefit programmes. Contemporary organisations have become more reliant on knowledge and the management of information. The present research findings will result in an increase in the stock of knowledge about pilot attitudes.

References

Allport, G.W. (1954). *The nature of prejudice*. Reading, Mass: Addison Wesley.

Australian Transport Safety Bureau (2005). *Aviation Safety Indicators: a report on safety indicators relating to Australian Aviation.* ACT: Australian Transport Safety Bureau.

Bateman. S. (1987). *"The Right Stuff" has no gender*. Aerospace Power Journal. Winter 1987-88. Retrieved March 19, 2004. (http://www.airpower.maxwell.af.mil/airchronicals/apj/apj87/bateman.html)

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PELIS, and SIMPLIS: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum.

Becker, B. E., Huselid, M. A., & Ulrich, D. (2001). *The HR Scorecard: Linking people, strategy and performance.* Boston: Harvard Business School Press.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: the case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*, 3-25.

Davey, C. L. (2004). The impact of human factors on Ab Initio pilot training. *Gender, Work and Organization, 11*(6), 627-647.

Davey, C. L. & Davidson, M. J. (2000). The right of passage: The experience of female pilots in commercial aviation. *Feminism and Psychology, 10*(2), 195-225.

Dockery, T. M. & Bedeian, A. G. (1989). "Attitudes versus action": LaPierre's (1934) classic study revisited. *Social Behavior and Personality, 17*, 9-16.

Glick, P. & Fiske, S. T. (2001). An ambivalent alliance: hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist, 56*, 109-118.

Helmreich, R. L. (2000). Culture and error in space: Implications from analog environments. *Aviation, Space, and Environmental Medicine, 71*(9, Section 2, Suppl.), A133-A139.

Helmreich, R.L & Merritt, A.C. (1998). *Culture at work in aviation and medicine.* Aldershot: Ashgate.

Helmreich, R. L & Merritt, A. C. (2000). Safety and error management: the role of Crew Resource management. In B.J. Haywood and A.R. Lowe (Eds.). *Aviation Resource Management* (pp. 107-119). Aldershot, UK: Ashgate.

Jöreskog, K. G. (1993). Testing structural equation models. In K.A. Bollen and J.S. Long (Eds.). *Testing structural equation models* (pp. 294-316). Newbury Park: Sage.

Karp, M., Turney, M. A., Green, M. F., Sitler, R. L. Bishop, J. & Niemczyk, M. (2002). Retaining women in collegiate aviation programs: Implementing learning style considerations. *International Journal of Applied Aviation Studies, 2*(1), 117- 131.

Merritt, A. C. & Helmreich, R. L. (1996). Human factors on the flight deck: the influence of national culture. *Journal of Cross-Cultural Psychology, 27*(1), 5-24.

Miller, L. & Hayward, R. (2006). New jobs, old occupational stereotypes: gender and jobs in the new economy. *Journal of Education and Work, 19*(1), 67-93.

Miller, N. & Brewer, M. B. (eds.). (1984). *Groups in contact: The psychology of desegregation.* New York: Academic Press.

Muchinsky, P. M. (2003). *Psychology applied to work: An introduction to industrial and organizational psychology*. (7th ed.). Belmont, CA: Thomson Wadsworth.

Oskamp, S. (1997). *Attitudes and opinions*. Englewood Cliffs, NJ: Prentice Hall.

Rollinson, D., Broadfield, A., & Edwards, D. J. (1998). *Organisational Behaviour and Analysis.* Singapore: Addison-Wesley.

Schein, V. E., Mueller, R., Lituchy, T. & Lui, J. (1996). Think manager – think male: a global phenomenon. *Journal of Organizational Behavior, 17*(1), 33-41.

Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics.* (3rd ed.). New York: Harper Collins.

Vermeulen, L.P., Wilson, J. & Mitchell, J.I. (2004) *The Measurement of Perceptions Regarding Gender-Related Pilot Behaviour: An exploratory study.* Paper presented at the 26th Conference of the European Association for Aviation Psychologists. Sesimbra: Portugal.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Mitigating the Loss of Navigational Awareness While Flying With GPS and Moving Map Displays Under VFR

Stephen M. Casner
National Aeronautics and Space Administration
Ames Research Center
Moffett Field, CA 94035-1000

Abstract

An earlier study demonstrated how reliance on GPS and moving map displays could significantly degrade pilot navigational awareness when flying under VFR (Casner, 2005). It was hypothesized that the drop in navigational awareness was due to the passive role assumed by pilots when using equipment that automates the navigation task. In this follow-up study, eight pilots used GPS and moving map displays to navigate between the same circuit of checkpoints used in Casner (2005) while performing one additional task: while en route between each pair of checkpoints, pilots were asked to choose and point out three geographical features. The research question was whether or not a greater involvement in the navigation task would result in better pilot performance on the same test of navigational awareness used in Casner (2005). Using the data from Casner (2005) as a control, a significant advantage was indicated for pilots who pointed out geographical features while navigating using GPS and moving maps. This suggests that simple practices that place the pilot in a more active role can help mitigate the "out-of-the-loop" phenomenon associated with using GPS and moving maps.

Introduction

Despite the many arguable advantages of using GPS and moving map displays, a previous study has shown how reliance on GPS and moving map displays can significantly degrade pilot navigational awareness (Casner, 2005). In that study, sixteen pilots were asked to fly, as accurately as possible, over a circuit of checkpoints in an unfamiliar area. Eight of the sixteen pilots were provided with a sectional chart (the Pilotage group). The eight remaining pilots were provided with the same sectional chart and a panel-mounted GPS receiver featuring

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.

a color moving map display (the GPS/Map group). Navigational accuracy was recorded at each checkpoint in the circuit. After navigating along the circuit, all pilots were unexpectedly asked to fly the same circuit again. This time, the Pilotage group was asked to navigate around the circuit without the use of the sectional chart, while the GPS/Map group was asked to navigate without either the chart or the GPS and moving map. Navigational accuracy was measured again for each checkpoint on this second trip around the circuit. The GPS/Map group performed significantly worse than the Pilotage group when navigation resources were taken away. Two pilots who used the GPS and the moving map were unable to find their way to the starting point of the circuit. Other GPS/Map pilots made large errors in navigating to individual checkpoints.

A simple depth-of-processing explanation (Craik & Lockhart, 1972; Glenberg, Smith, & Green, 1977) was offered for the degraded performance among pilots who used GPS and moving maps. Pilots who used only the sectional chart for navigation were forced to take careful note of geographical features and actively use them to locate checkpoints. This navigational method required deep processing of geographical features and resulted in a high degree of familiarity with the area. Pilots who relied only on the GPS and moving map were free to set aside the sectional chart and largely ignore geographical features as they were automatically guided to each waypoint by the GPS computer. When confronted with a situation in which familiarity of the area was suddenly needed, pilots who were actively engaged in the navigation task performed well. Pilots who relied on GPS and who did not actively participate in the navigation process performed poorly. Endsley (1996) cited a number of studies in which a similar effect has been demonstrated when human operators are combined with automated systems.

Mitigating the Negative Effects of GPS and Moving Maps

Given the many advantages of GPS and moving maps (e.g., locating the nearest airport during an emergency), it is difficult to argue that pilots should not use them. A more sensible approach is to ask: Are there simple practices that pilots can adopt that allow them to take advantage of the beneficial features of GPS and moving maps, yet avoid the “out-of-the-loop” phenomenon?

In this study, a third group of eight pilots was asked to navigate around the same circuit of checkpoints using the same GPS receiver, moving map display, and sectional chart. This group of pilots was asked to perform one additional task while making their way around the circuit of checkpoints. The experimenter asked each pilot to choose and point out any three geographical features of interest between each pair of checkpoints in the circuit – a total of fifteen geographical features for the entire circuit of checkpoints. It was explained to each pilot that the purpose of this task was to prevent the pilot and experimenter from becoming bored during the flight. The pilot did not need to possess or look up any information about the geographical features, just simply choose and point out interesting-looking features along the way.

In terms of the deep vs. shallow processing hypothesis, pilots who point out geographical features represent a middle ground: these pilots are neither wholly burdened with the navigation task, nor wholly excused from it. If we compare the performance of this third group of pilots to the performance of the two groups from the earlier study, a number of questions can be answered. Is the cognitive effort required to choose and point out geographical features sufficient to avoid the out-of-the-loop phenomenon observed among users of GPS and moving maps? How does the navigational awareness of these pilots compare to that of pilots who navigate using more labor-intensive pilotage methods? How does their awareness compare to that of pilots who relied solely on GPS? Can the practice of pointing out geographical features serve as a practical technique for VFR pilots who use GPS and moving maps?

Method

Participants

The same criteria used in the previous study were used to recruit additional eight pilots. All pilots were legally qualified to act as pilot in command in the experiment airplane. All pilots had basic familiarity with GPS receivers and moving maps. All pilots reported that they did not have significant familiarity or experience with the area in which the data were to be collected (Casner, 2005).

Apparatus

The same Diamond DA40 (Diamond Star) equipped with a panel-mounted GPS receiver and a color moving map display was used for data collection. All pilots were given a current San Francisco sectional aeronautical chart that covered the area through which the experimental flight was conducted. The experimenter used an additional GPS receiver, hidden from pilots' view, to measure navigational accuracy (Casner, 2005).

Procedure

As with the earlier study, the data were collected in Northern California, during July and August, under VFR conditions with a reported visibility of greater than six statute miles (P6SM) at all nearby airports. Prior to engine start, the eight pilots were given a briefing similar to that given to the pilots from the earlier study (Casner, 2005). Pilots were told that the flight would require them to navigate along a series of nine cross-country checkpoints. A sectional aeronautical chart was used to point out each of the checkpoints. Pilots were told that the first three checkpoints were to be considered practice checkpoints, and that the last six checkpoints, shown in Figure 1, were the ones of interest to the experimenter.

Pilots were instructed to fly over each checkpoint as accurately as possible, and to report when they believed that they were directly over each checkpoint. Pilots were free to choose altitudes appropriate for VFR flight at their discretion.

All eight pilots had available a sectional chart and a GPS with a color moving map display. The experimenter confirmed that each pilot was familiar with the basic features of the GPS and moving map prior to departure. The series of nine checkpoints was programmed into the GPS prior to takeoff.

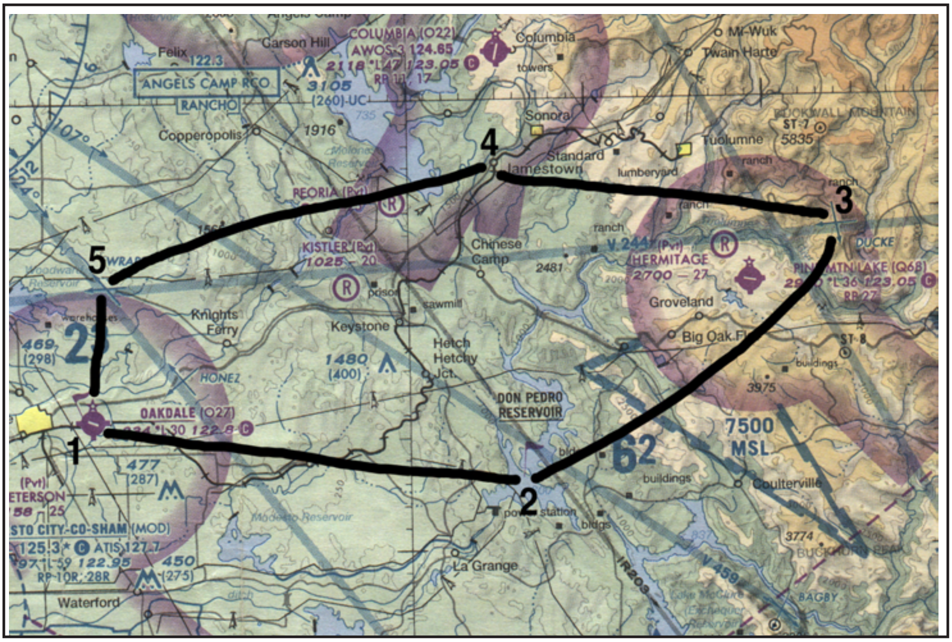


Figure 1. Sectional chart showing the circuit of six checkpoints used in the present and earlier study (Casner, 2005).

En route to each checkpoint, pilots were asked to choose and point out three geographical features of interest. Pilots were told that they did not have to know anything about the geographical features they pointed out, or look up any information about them.

As pilots reported reaching each checkpoint, the experimenter used a second GPS receiver, hidden from the pilot's view, to record the true distance from the checkpoint.

After completing the circuit of six checkpoints shown in Figure 1, the experimenter took away the sectional chart, turned off the GPS and moving map, and (unexpectedly) asked each pilot to fly the circuit of six checkpoints again.

The eight pilots flew over the loop of six checkpoints once again, reported crossing each checkpoint, while the experimenter again noted the navigational error at each checkpoint.

At the conclusion of the flight, pilots were debriefed on the purpose of the study. The importance of remaining actively involved in the navigational process was emphasized with all pilots.

Results

The purpose of the present study was to measure the effect of pointing out geographical features of interest on navigational awareness among users of GPS and moving maps. For this reason, the results for this group of pilots are compared to the two groups from Casner (2005). Thus, the analyses below present data for three groups:

1. Pilotage: Pilots who used sectional charts only [from Casner, 2005];
2. GPS/Map: Pilots who used sectional charts, GPS, and moving maps [from Casner, 2005];
3. GPS/Map with Callouts: Pilots who used sectional charts, GPS, moving maps, and pointed out geographical features of interest.

Navigation Error: First Pass

The graph in Figure 2 shows the mean navigational errors during the first pass through the checkpoints for the three groups of pilots.

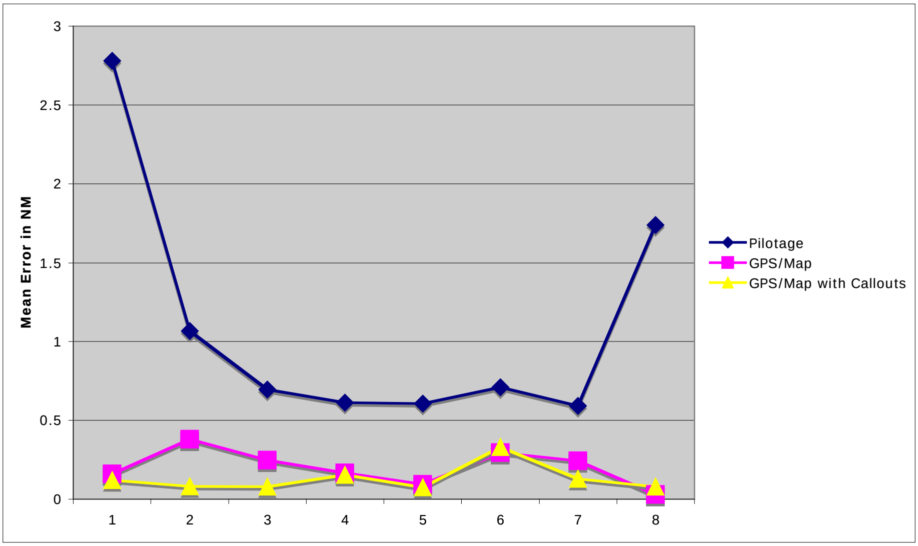


Figure 2. Navigational accuracy with all navigational resources available.

The mean navigational error and standard deviation for the three groups were: Pilotage = 1.1 NM (1.5 NM); GPS/Map = 0.2 NM (0.3 NM); and GPS/Map with Callouts = 0.13 NM (0.7 NM).

During the first pass through the circuit, with all navigational resources available, the group that pointed out geographical features was statistically indistinguishable from the GPS/Map group in the previous study that did not point out geographical features. The GPS/Map with Callouts group performed as well as the GPS/Map group, and significantly better than the Pilotage group ($t = 3.48$, $p < 0.01$), although all three groups performed within the 3 NM navigation standard for pilotage and dead reckoning cited in the Private Pilot Practical Test Standard (FAA, 2002).

Navigation Error: Second Pass

The graph in Figure 3 shows the mean navigational errors during the second pass through the circuit for all three groups: when pilots had all navigation resources taken away from them.

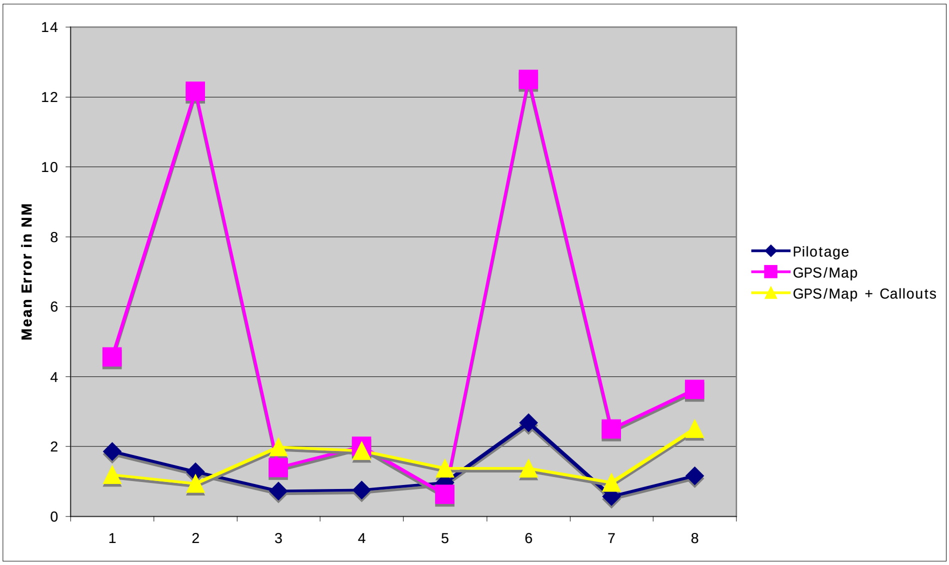


Figure 3. The mean navigational errors during the second pass through

The data in Figure 3 show that the practice of choosing and pointing out geographical features resulted in a significant improvement in navigational performance for users of GPS and moving maps. While the mean navigational error for the GPS/Map group was 4.92 NM (7.92 NM), navigational error for the GPS/Map group that pointed out geographical features was 1.53 NM (1.42 NM).

Figure 4 summarizes, in a single graph, the navigational performance of all three groups with and without navigational resources available. Indeed, it appears that the simple task of choosing and pointing out geographical features significantly lessens the "out-of-the-loop" effect suffered by GPS and moving map users.

As with the two groups of pilots from the previous study, the eight pilots recruited from the present study varied widely in their total flight experience [min = 160 hours; max = 8800 hours; mean = 1968 hours; median = 815 hours]. There were no significant differences for total flight time between any of the three groups compared here, or significant correlation between flight time and navigational performance.

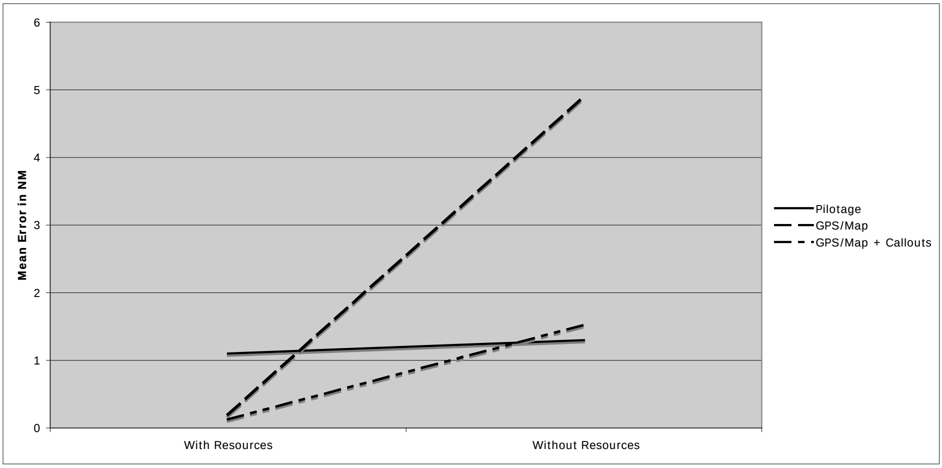


Figure 4. Navigational performance of all three groups with and without navigational resources available

Conclusion

The data show that pilots who use GPS and moving maps, and who invest the time to take note of geographical features along their route of flight, exhibit a level of navigational awareness that is higher than pilots who make no such effort. This finding suggests two things: (1) there are practical techniques that can help mitigate the loss-of-awareness phenomenon observed among pilots who use GPS and moving maps; and (2) a more active pilot involvement in the navigation task seems to be the key to maintaining navigational awareness. What is perhaps most interesting about the result is how such a simple practice of pointing out geographical features was sufficient to make such a striking difference in pilot awareness. This suggests that navigational awareness is indeed a fragile phenomenon.

While it is tempting to conclude that the simple technique of pointing out geographical features represents the solution to the loss-of-awareness problem, we must refrain from doing so for a number of reasons. First, only a small sample of pilots was tested (eight pilots per group). Two of the eight pilots who passively used the GPS and moving map got lost. While it is fair to suggest that the practice of pointing out geographical features lowers the likelihood of getting lost to something less than one-in-four, it is an open question of what would happen if a hundred or a thousand pilots were to complete the study. Second, the measure of navigational awareness used for the study is far from comprehensive. In all conditions, pilots were asked to perform the relatively straightforward task of navigating along the same route a second time. Thorndyke and Hayes-Roth (1982) nicely demonstrate the difference between acquiring knowledge required to replicate a route and acquiring the knowledge required to solve more generalized navigation problems such as finding one's way to a different destination, or finding a different route to the same destination. To what extent the familiarity gained by pilots who pointed out geographical features would serve to solve more complex (and realistic) navigational problems deserves future study. Third, the technique of pointing out geographical features is simply not possible in all situations. For example, it is generally not possible to see geographical features when flying in instrument meteorological conditions. Even under visual meteorological conditions, other cockpit duties (e.g., scanning for traffic, configuring avionics, etc.) would often prevent pilots from performing an out-the-window search for geographical features. Hence, there is a need to discover other practical techniques that help pilots maintain navigational awareness.

A future study might systematically consider what kinds of involvement in the navigation task serve to keep pilots in the loop. Parasuraman (1996) reviews a number of studies that explore different techniques for sharing duties between human operator and automated system. In addition, working with context-rich information such as geographical features yields results that are different from working with more abstract information such as bearings and distances. A better understanding of these factors might contribute to the design of effective practices for maintaining awareness.

The results reiterate the distinction between navigational awareness existing in the storage registers of a computer and navigational awareness actively circulating in the head of the pilot. Casner (2005) demonstrated the consequences for the case in which the GPS and moving map become inoperative or unavailable. Riley (1996) reviews a number of problems that can occur when human operator and automated system have differing assessments of a task in progress – when both entities are operational. Riley identified a number of factors that can cause human operators to disregard the indications of an automated system in favor of their own mistaken beliefs, or disregard their own accurate beliefs in favor of the erroneous indications of an automated system. These findings suggest that, as long as the task of navigating an aircraft is shared between human operator and automated system, it is not acceptable to place all of the responsibility for maintaining navigation awareness on a GPS receiver or similar device. Pilots, flight instructors, evaluators, and policymakers have long talked about the importance of "staying in the loop" while flying with automation. Perhaps now is a good time to make explicit proficiency standards for navigational awareness in the technically advanced cockpit.

References

Casner, S. M. (2005). The effect of GPS and moving maps displays on navigational awareness while flying under VFR. International Journal of Applied Aviation Studies 5 (1), 153-165.

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

Endsley, M. R. (1996). Automation and situation awareness. In Parasuraman, R., and Mouloua, M. (Eds.), Automation and Human Performance, (pp. 163-181). NJ: Erlbaum Associates.

Federal Aviation Administration (2002). Private pilot practical test standards for airplane. FAA-S-8081-14A.

Glenberg, A. M., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. Journal of Verbal Learning and Verbal Behavior, 16, 339-352.

Parasuraman, R. (1996). Monitoring of automated systems. In Parasuraman, R., and Mouloua, M. (Eds.), Automation and human performance, (pp. 91-115). NJ: Erlbaum Associates.

Riley, V. (1996). Operator reliance on automation: Theory and data. In Parasuraman, R., and Mouloua, M. (Eds.), Automation and Human Performance, (pp. 19-35). NJ: Erlbaum Associates.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. Cognitive Psychology 14, 560-589.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Introduction of Technically Advanced Aircraft in Ab-Initio Flight Training

Nicklas Dahlström and Sidney Dekker
Lund University School of Aviation
Ljungbyhed, Sweden

and

Staffan Nählinder
Swedish Defence Research Agency
Linköping, Sweden

Abstract

The transition of pilots from a traditional cockpit to a modern glass-cockpit has been a training challenge for the last two decades. The arrival of Technically Advanced Aircraft (TAA) during the last decade has brought the opportunity to introduce this technology from the beginning of airline pilot training. In this project, three flight instructors responsible for the introduction of TAAs in ab-initio training at a flight school were interviewed on their initial experiences and concerns regarding the introduction. Subsequently, questionnaires were collected from the familiarization training of instructors on the new aircraft and from ab-initio students and instructors after three of the 18 flights leading up to the first solo. Finally, flight instructors involved in the introduction were interviewed. The results showed that anticipated problems with use of displays, aircraft speed, and use of side control proved to have limited impact on the training. The conclusion is that with extensive preparation, introduction of TAA in ab-initio training can be accomplished successfully. However, the expected benefits of this on training and questions on what might be lost in the process need to be addressed by further research.

Introduction

The most important training challenge in commercial aviation since the eighties has been the training of pilots transitioning from a traditional cockpit environment to that of a modern computerized cockpit with glass-cockpit and sophis-

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.

ticated automation (Billings 1997; Dekker & Hollnagel, 1999). The transitions produced unanticipated situations and reactions, sometimes resulting in incidents and accidents, and proved to the aviation industry that such a technology shift transforms work in the cockpit and cannot be treated as a separate subject or an add-on to existing training (Rignér & Dekker, 1999). Although the airline industry has invested in increasing the effectiveness of training for the modern cockpit, training new pilots for licensing has not changed at the same pace (Dekker & Johansson, 2000).

The transition to a modern cockpit environment often occurs late in the training of a new pilot, usually concurrent with the pilot being introduced to multi-crew and jet-transport flying. To avoid or alleviate problems with this transition Rignér & Dekker (1999), as well as Casner (2003a; 2003b), recommended that learning about the glass-cockpit should be introduced at early stages of pre-airline pilot training. Due to the restricted funding at collegiate or private level flight training this has often been difficult (Fanjoy & Young, 2003), with the consequence that there has been large variations in how well prepared for the modern cockpit new pilots have been when they enter an airline.

The arrival of modern cockpit technology in light general aviation aircraft seen in the last decade has provided the opportunity to introduce, from the beginning, a modern cockpit environment to pre-airline training. These aircraft, known as Technically Advanced Aircraft or TAA, are equipped with most of the technology found in large transport aircraft, except for a Flight Management System. According to a Federal Aviation Administration (2003, p. 9) report, a TAA is an aircraft, “in which the pilot interfaces with one or more computers in order to aviate, navigate, or communicate.” A more strict definition is also provided:

A TAA is defined as an aircraft that has at a minimum:

a. IFR –certified GPS navigation equipment (navigator) with moving map; or
b. A multi-function display (MFD) with weather, traffic or terrain graphics
c. An integrated autopilot.

According to a report by the AOPA Air Safety Foundation (2005), many new TAAs go beyond this definition, also including a Primary Flight Display (PFD) to replace the traditional “six-pack,” and coming close to the glass cockpit concept large transport aircraft. According to the same report, “Fleet sales to active flight schools and university flight departments in the last two years have generally been TAA” (p. 4).

Introducing TAAs in pre-airline training is a way to reproduce the technology shift that has already taken place in air transportation. The benefits of this seem obvious; the cockpit of a TAA resembles the future work environment and introduces the type of instrumentation and automated functions used in jet transport aircraft. However, since a technology shift transforms work, it is also likely to transform the learning of students and the teaching of flight instructors. In the previously mentioned report of the AOPA Air Safety Foundation (2005, p. 22), executive director Bruce Landsberg of the organization stated, “Technology emerges as a doubled-edged sword, increasing pilot and aircraft capabilities but frequently at the price of increased workload and education."

Since the introduction of advanced computer technology in the cockpit there has been research on how to effectively train for this environment (Roessingh et al., 1998; Dekker & Hollnagel, 1999). There also has been research regarding how to train student pilots for the modern cockpit. Fanjoy and Young (2003; 2004; 2005) studied collegiate flight training programs to see how glass-cockpit training was addressed in these programs. One of their conclusions was, although a vast majority of the flight training programs that they studied seemed to share the belief that flight automation training was critical to the success of their students, few applied comprehensive training in that area. Casner (2003a; 2003b) studied classroom training for cockpit automation and the transition from piston trainer aircraft to jet transport aircraft, using a computer based simulator of the flight management system of a modern jet. Casner concluded that "cockpit automation found in small training airplanes appears to provide a simple, cost-effective way of introducing cockpit automation to pilots who are still in the formative phases of their professional aviation careers" (2003 b, p.16). Craig, Bertrand, Dornan, Gosset, and Thorsby (2005) compared the use of TAAs, an adapted syllabus, and scenario-based flight exercises for training student pilots with traditional training. The first data from this project showed that students using TAAs had to repeat more (61% vs. 17%) flight exercises than those in traditional training before the first solo, but had to repeat less flight exercises during private pilot and cross-country phases of training (15% vs. 38%) and during instrument training (24% vs. 45%).

The aim of this research project was to study the extent and nature of the transformation of pre-airline pilot training as an effect of the introduction of TAAs. This included investigating the perceived, as well as the actual, benefits and problems of training, in addition to the identification and monitoring of potential flight safety risks. The project focused specifically on the first phase of the introduction, i.e. the training phase up to the first solo flight, since it was expected that this phase would represent the "leap" in the introduction of the new technology and as such would require the greatest efforts from students and flight instructors with adapting to it. However, the aircraft is planned for the flight training phases prior to instrument training, which then will be performed in a traditional twin piston-engine aircraft. Even though TAAs are increasingly being purchased by flight schools and university flight departments, a majority of these organizations still have an introduction of TAAs ahead of them. Thus, the findings of this project provide valuable contributions not only to the new field of knowledge in general aviation and aviation training represented by TAAs, but also to flight training organizations planning to introduce TAAs into their training.

Method

Participants

The participants were flight instructors and students at Lund University School of Aviation. The 17 flight instructors were all male and between 26 and 62 years of age. Their flight experience ranged from 470 to 28,500 flight hours and their instructor experience from 150 to 14,500 hours. The 12 ab-initio students were at the start of their 20-month Integrated Airline Transport Pilot training program. The students were 20 to 30 years old and three out of twelve were female. Two were holders of a Private Pilot License (PPL), with 139 and 67 hours of experience respectively, one had a license for touring motor glider and 100 hours of experience, while the other students had minimal or no flying experience.

Aircraft

The aircraft to be introduced in the training was the Cirrus SR-20 G2. The aircraft is a TAA well beyond the FAA definition, with equipment including PFD, MFD (with moving map and traffic warning system), GPS, and advanced autopilot functions. It has a side control instead of a traditional yoke and a single power lever instead of the traditional two levers for throttle and propeller. A parachute is integrated in the airframe and designed to be used when a controlled landing is not an alternative. Higdon (2000, ¶ 1) stated in a review of the aircraft that it "stands out because of its size, its comfort, its equipment" and that it was "the best-handling, best harmonized flying machine since the Bonanza" (¶ 3). Yet Higdon was negative about using it for students and predicted that "getting the hang of basic airmanship while trying to manage this extremely slippery bird makes for more work than most students need to face early in their training" (¶ 4).

Procedure

Interviews with flight instructors prior to the introduction of the new aircraft. Semi-structured interviews with three flight instructors were performed to collect initial experiences of the aircraft and to map areas of concern. Those interviewed were the flight instructor responsible for the introduction, the chief flight instructor, and a senior flight instructor. They had been selected to prepare the introduction, i.e. to fly the aircraft to learn about its handling and performance, revise and adapt the training and training material, and to perform familiarization training on the new aircraft with other instructors. They were also included in a group of six flight instructors designated for the first course with the aircraft.

Questionnaires after familiarization training for flight instructors. Familiarization training together with a flight instructor trained on a new type of aircraft is not mandatory for flight instructors. According to the regulations, flight instructors are authorized with performing their own familiarization training. However, this familiarization training intended to give all flight instructors in the organization a well-calibrated and standardized first acquaintance with the aircraft. After receiving familiarization training, the experiences of 14 instructors were collected with a questionnaire. The questionnaire was based on the areas of concern brought up in the previous interviews and designed in cooperation with the flight instructor responsible for the introduction of the aircraft.

Questionnaires for students and flight instructors during flights up to first solo. A questionnaire was also used to collect the experiences of the students and instructors for the flights up to the first solo. It was based on available information from interviews and the previous questionnaire and designed and revised between flights in cooperation with the flight instructor responsible for the introduction. Three flights out of the 18 leading up to the first solo were selected. The first flight chosen for the questionnaire was exercise 104, the fourth flight for the students. Level turning was practiced in this flight and it was chosen since it represented the first flight where the students themselves handle the aircraft most of the time in the air. In the second selected flight, flight exercise 111, power-off landings were practiced. Finally, the third flight exercise, flight 118, contained practice of take off and landings as preparation of the upcoming first solo. Repeating the same questions provides stable answers that facilitate generalization beyond the particular flights and provides opportunities to see trends of learning.

Interviews with flight instructors after introduction of the new aircraft. To complement questionnaire data three semi-structured interviews were performed after the first solo, including following up on flights after the first solo and the skill test for PPL. Since the PPL skill test was performed by external examiners, it was expected to provide a calibration of the views of the instructors on the introduction of the new aircraft.

The flight instructor responsible for the introduction was interviewed again since he had monitored the whole introduction. The second instructor was the course manager, who was expected to have an overview of the experiences of students and instructors. The third instructor was the flight safety pilot, who also was an instructor on this course. The two instructors interviewed prior to the introduction were not interviewed again due to potential bias towards a successful perception of the introduction.

Results

Interviews with flight instructors prior to the introduction of the new aircraft

The three instructors spent 15 flight hours and ample time on the ground during four months to prepare the introduction of the new aircraft. The initial experiences of the aircraft brought up five areas of concern; the computer-driven instrumentation (PFD and MFD), the speed of the aircraft, the use of side control, the work environment and the safety of the aircraft.

Adapting instrument scanning to the PFD included expected problems of not finding the right information at the right moment and specific problems with the presentation of speed and altitude on tapes. The altitude and speed of the aircraft is highlighted in a box on the tape. The consequence is that as numbers roll on the tapes they are partly covered by the box and, especially at low speeds, this was considered a problem. The increased precision provided by the tape presentation was another problem, since this created excessive focus on the numbers on the screen. One instructor commented that when flying at 1500 feet with a traditional altimeter he knows that he is at about the right altitude. Seeing on the PFD that he is flying at 1480 feet made him wonder why he could not stay at the correct altitude. The access to large amounts of information on the MFD was seen as potentially detrimental to the attention and mental workload of the students. Due to this, it was decided that prior to the first solo the students should only be allowed to have engine information presented on the MFD (unless other information would be needed for flight safety reasons). Due in part to the same reasons use of automation was also planned to be restricted until later stages of training.

The new aircraft, like many TAAs, operates in a higher speed range than traditional single piston-engine trainer aircraft and the instructors anticipated that this would have consequences for the training. Consequently, power settings and speeds recommended by the manufacturer frequently had to be reduced, approach points had to be changed, and flight profiles modified. Still, the effect of the higher speed range on the training of new students was a concern. The side control was experienced as unproblematic by the instructors, besides concerns that it would strain muscles in the arm during intense maneuvering. The aircraft frequently needs trimming in the roll-axis and with the trim-button being used for both roll and pitch, it was considered that accidental trimming in the pitch-axis could occur.

The flight instructors were convinced that the improved work environment (space, seats and headsets with active noise reduction) would lead to an improved learning situation and more effective preparation of the students for modern jet transport aircraft. Absence of dual instrumentation, a feature of the previous aircraft, did not seem to be of concern to the instructors. Looking at the PFD and MFD and monitoring the use of the side control from the right seat was considered unproblematic. The information available on the MFD was considered to improve safety and could be used as back-up on cross-country flights if a student would get lost. The parachute of the aircraft did not seem to be of importance for the perception of safety. Contrary, one instructor brought up the risk of such a device resulting in over-confidence, the problem of knowing when to deploy the parachute, and argued that a crash with a parachute might be no better than a controlled emergency landing.

Questionnaires after familiarization training for flight instructors

The questionnaire was returned by all of the 14 flight instructors going through the familiarization training. The training consisted of two flights, handling of the aircraft (maneuvering at slow speed, stall and recovery, traffic circuits, and landings) and a cross-country flight (to practice use of the PFD, MFD, and GPS-panel). For the questionnaire, a Likert scale from 1 to 9 was used, with different labels selected for the end points of the scale (Nählinder, Berggren & Persson, 2005) and the instructors were requested to provide comments to the questions. The results are shown in table 1 and commented below.

Table 1

Questionnaire results after flight instructor familiarization training on new aircraft

Questions	Mean (Std.dev.)
Preparation for training (not well-very well prepared)	4.6 (2.2)
Difficulty of training (very simple-very difficult)	3.4 (1.3)
Performance during training (not good-very good)	6.7 (1.2)
Time pressure during training (none-great)	2.6 (1.8)
Time spent looking at PFD/MFD (little-plenty)	6.7 (1.1)
Effect of PFD/MFD on performance (worse-better)	4.9 (1.4)
Misunderstanding/confusion due to PFD/MFD(none-often)	3.9 (1.7)

Disturbed by other information on PFD/MFD (none-often)	3.8 (1.6)
Experience of use of tape presentation (not good-good)	5.1 (2.2)
Effect of precision on PFD on performance (worse-none)	5.0 (2.7)
Perception of speed (low-high)	5.1 (1.5)
Effect of speed on performance (worse-none)	8.2 (0.8)
Maneuvering with side control (not good-good)	7.4 (1.7)
Difficulties with side control (plenty-none)	7.0 (1.8)
Handling of other controls (simple-difficult)	2.6 (1.1)
Extent of using automation (little-plenty)	5.0 (2.6)
Handling of automation (simple-difficult)	3.2 (1.8)
Work environment of aircraft (not good-very good)	8.2 (0.8)
Projected success of aircraft in flight training (not good-very good)	7.3 (1.3)
Familiarization training (not sufficient-sufficient)	6.1 (2.6)
Safety compared to previous aircraft (less-more safe)	4.9 (1.8)

Eight instructors commented on problems of finding the right information at the right moment on the PFD/MFD, with one stating that he happened to read altitude as vertical speed in feet. Five comments on the speed all stated that it was not a problem. The side control was commented by nine instructors: four about potential strain on muscle of the arm, three on the aircraft being sensitive in the roll-axis and two on the need for trimming. Other comments on controls brought up problems with closing the door and difficulties with finding or handling the Emergency Locator Transmitter (ELT), brakes, parking brakes, circuit breakers, alternate warm air, and alternate static air. Comments on the work environment included two concerns on the strength of the construction and one on the aircraft being sensitive to judgment errors regarding speed and attitude when landing. In addition, one instructor stated that students would become less skilled with flying and more skilled with information management and automation. Another stated that he was not sure if training with the new aircraft would cover the knowledge needed when flying an aircraft at a flying club, landing on grass strips or navigating in poor visibility with traditional instruments. The rating of the aircraft as more or less safe compared to the previously used aircraft was neutral. Safety risks were commented by twelve instructors; four on stall-properties of the aircraft, two on the risk for stall in the touch-down phase, two on the risk of looking too much on the instruments and one on the PFD/MFD not being stable and having to be restarted often.

Questionnaires for students and flight instructors during flights up to first solo

From 35 flights, 33 questionnaires were returned from the students and 34 from the instructors. One student holding a PPL did not fly flight 118, which led to the loss of one student and flight instructor questionnaire respectively. One student did not return two questionnaires; one instructor did not return one questionnaire. The questionnaires used the same scale as that for the familiarization flights and comments from both students and instructors were requested. After the first flight, questions on changes between flights were added and questions on the PFD/MFD separated.

Table 2
Results from questionnaires after three selected flights, means and standard deviations for ratings of students and instructors (shaded).

Questions	Flight 104	Flight 111	Flight 118
Preparation (not well-well prepared)	7.6 (0.8)	8.0 (1.1)	7.7 (1.7)
	7.6 (1.0)	7.7 (1.0)	7.6 (1.1)
Difficulty (very simple-very difficult)	3.9 (1.5)	6.1 (1.5)	4.8 (2.1)
	5.5 (1.7)	5.6 (1.6)	5.6 (1.7)
Performance (not good-very good)	7.5 (0.9)	6.7 (1.1)	7.3 (1.1)
	7.1 (0.9)	6.3 (1.4)	6.7 (1.0)
Time pressure (none-great)	2.1 (1.6)	2.7 (1.8)	2.4 (1.5)
	3.3 (1.5)	4.0 (1.8)	2.9 (1.3)
Spare time (none-plenty)	4.6 (2.6)	2.4 (2.3)	3.1 (2.5)
Step-by-step or automated actions	5.6 (2.7)	5.8 (2.0)	6.5 (1.6)
	6.2 (1.5)	5.2 (1.8)	6.8 (1.5)
Forced to “shut off” information (never-often)	2.2 (1.2)	2.4 (1.7)	2.1 (0.9)
	2.9 (1.4)	3.1 (2.0)	1.7 (1.0)
Forced to interrupt other tasks to maneuver the aircraft (never-often)	1.5 (0.7)	1.2 (0.5)	1.2 (0.6)
	1.5 (0.9)	1.9 (1.5)	1.5 (0.8)
Use of PFD (less-more)	-	4.5 (2.2)	4.8 (1.8)
	-	4.4 (1.6)	4.6 (1.0)
Misunderstanding/confusion PFD (none-often)	2.5 (1.8)	1.8 (1.4)	2.0 (1.9)
	2.4 (1.4)	2.2 (1.7)	1.7 (1.2)

Disturbed by information on PFD (none-often)	2.5 (1.8)	1.5 (1.0)	1.4 (0.7)
	2.8 (1.7)	2.2 (1.6)	1.5 (0.5)
Use of MFD (less-more)	-	4.8 (1.9)	4.8 (1.9)
	-	4.6 (1.8)	4.9 (0.7)
Misunderstanding/confusion MFD (none-often)	2.5 (1.8)	1.8 (1.3)	2.5 (1.9)
	2.4 (1.4)	1.8 (1.5)	1.7 (1.2)
Disturbed by information on MFD (none-often)	2.5 (1.8)	2.0 (1.8)	2.1 (2.0)
	2.8 (1.7)	1.5 (1.7)	1.6(1.2)
Perception of speed (low-high)	4.8 (1.8)	5.2 (1.0)	5.1 (0.6)
Effect of speed on performance (worse-none)	6.8 (1.9)	7.2 (1.5)	7.6 (1.5)
	8.0 (1.5)	8.0 (1.8)	8.5 (1.0)
Maneuvering with side control (not good-good)	7.2 (1.4)	7.6 (1.4)	7.7 (1.4)
	8.2 (1.1)	8.4 (1.2)	8.6 (0.8)
Difficulties with side control (plenty-none)	6.6 (1.2)	7.5 (2.0)	7.4 (1.9)
	8.1 (1.3)	8.1 (1.4)	8.8 (0.4)
Handling of other controls (simple-difficult)	2.5 (1.6)	2.7 (1.9)	1.9 (0.9)
	2.4 (0.9)	1.7 (0.8)	1.4 (0.5)
Work environment (not good-very good)	8.2 (1.3)	8.2 (1.0)	8.0 (1.5)
Effect of no dual instr. (worse-none)	8.5 (0.6)	8.5 (0.8)	7.9 (1.6)

A repeated-measures analysis of variance showed a within-subject significant difference in the ratings of “Difficulty” and “Performance” (across the three flights) by both students and instructors, with flight 111 rated higher on “Difficulty” ($F[2, 38]=3.57$; $p<.05$) and lower on “Performance” ($F[2, 38]=4.75$; $p<.05$). Ratings on the category “Disturbed by information on the PFD” dropped significantly ($F[2, 36]=4.02$; $p<.05$) after the first flight. Maneuvering with side control was rated higher for successive flights and difficulty lower. Even though not significant at the 5% level, this tendency is clear ($F[2, 38]=2.67$; $p<.10$, and $F[2, 38]=3.18$; $p<.10$, respectively).

Flight 104 resulted in six comments from four students on problems regarding the presentation of information, two on difficulties with finding information, one on the precision of the altitude presentation, one on the speed presentation as confusing, and two that were positive to the PFD. Four of the instructors commented on the use of PFD as unproblematic, while one commented on excessive focus on the displays and one that the student had monitored Ground Speed (GS) and True Air speed (TAS) rather than Indicated Air Speed (IAS). One instructor commented that use of side control became strenuous due to intensive maneuvering and one student commented that the trim made him slant right. Flight 111 resulted in four students commenting positively on the PFD, although one student stated that he likes traditional instruments better. The speed was commented by seven students as unproblematic, although one brought up the difference between start and landing speed and one the problem of reducing the speed, i.e. the aircraft being "slippery". Flight 118 received few comments, one on improvements in finding information on the PFD, one on preference of traditional instruments, three positive comments on speed, and one bringing up the problem of speed reduction.

Interviews with flight instructors after introduction of the new aircraft

The three instructors unanimously expressed that the introduction had been less problematic than they had expected it to be. According to the course manager 3 flights had to be re-flown out of approximately 18 flights each for 12 students, all of them take-off and landing flights just before the first solo. (Approximately 18 since some flights were cancelled for the two students who already had a private pilot license.) The highlighting of numbers and precision of the PFD turned out not to be a problem at all for the students and a passing one for instructors. The course manager stated that in the beginning the amount of information on the PFD was beyond the capacity of the students, referring to students who had monitored the wrong speed on the PDF. The speed of the aircraft turned out to be less of a problem than expected. Two instructors commented that the aircraft is sensitive before touch-down; a marginal speed decrease below the landing speed necessitates the instructor to intervene. Three confirmed tail strikes in the initial stages of the training, although none causing significant damage had confirmed this problem. (These events raised the awareness of the problem among instructors and subsequently no tail strikes occurred.) The course manager stated that the aircraft is "slippery"; particularly during descent, speed reductions are difficult. According to the course manager there had been initial skepticism expressed towards the side control and its construction. However, the instructors all agreed that the use of the side control and other controls had been unproblematic. While initial concern had been expressed on the difficulties of recognizing when the aircraft is about to stall this also proved not to be a problem. The work environment of the aircraft was collectively praised as providing an improved learning environment. Regarding safety, the issue of potential over-confidence due to the information available on the MFD, primarily the moving map, was brought up.

The planning and preparation of the training and training material (particularly the hints for instructors) were unanimously seen as the main reason for the successful introduction. The instructor responsible for the introduction emphasized detailed planning of flight exercises to the pace of students' learning as a significant contributing factor. Considered as important was that a group of six flight instructors were designated and given time to prepare for the course. Preparations continued with instructor meetings before and after each flight and opportunities to try out flight exercises before they were performed with students. The documentation of meetings and practice flights in the form of hints for instructors was considered of great importance. The instructors quickly became well coordinated and calibrated, adding confidence to the instructors and avoiding otherwise common frustrations of students experiencing different training with different instructors.

Among the more recent experiences mentioned, was an event immediately after the first solo, when a student engaged the flight automation and then was unable to disengage it. After the PPL examination flights, the examiners commented on the students as being highly skilled on instrument flight but overall, they were overly focused on the instruments. Their knowledge of technical systems was considered below the normal standards of the flight school and their awareness of fuel planning and management even more so.

Discussion

The results of this study showed that the introduction of TAAs in ab-initio training can be accomplished successfully, with few problems for students and flight instructors. That three students needed to have an extra flight with touch and go training before the first solo is normal at this flight school and indicated that the progress of this course has not deviated significantly from that of previous courses. Comments showed that the concerns during the preparations for the introduction did translate into problems for some students, but never for a majority of the students. Credit for the successful introduction belongs to the preparation that took time and resources beyond the regulatory requirements (preparatory flights to try out aircraft, revision of training material, coordinated familiarization training, and allocating a selected group of instructors with time for preparing and following up). New technology often promises increased capabilities (for training: same as before, but now simpler and better) and safety, but the transformation of work changes how the performance breaks down and creates possibilities for new forms of accidents (Dekker, 2004). The promise of new technology applied to the introduction of TAAs in training implicates that only minimal preparations should be necessary. However, in its report, the Federal Aviation Administration (2003, p. 6) concluded that TAAs provide “potential for increased safety,” but to achieve it, additional training of specific TAA systems is necessary. The time and resources available for this introduction seem to have made it possible to prepare for the transformation of work and to avoid the initial pitfalls of the shift in technology.

The ratings of both students and instructors suggested that difficulties with learning how to operate the aircraft were not connected to its specific properties as a TAA; while flight 111 proved more difficult than other flights, ratings of use of PFD/MFD, speed and side control decreased or remained at the same level compared to those of flight 104. The concerns for the PFD identified in the first set of interviews (finding information, indication of altitude and focus on instruments) were replicated in the comments of the students and instructors after flight 104 but disappeared with increasing experience. Ratings on these problems remained low from the first flight to the last. These initial problems suggested that familiarization training on a part-task trainer or on a computer could be helpful in order to facilitate the use of the PFD and MFD. The higher speed range became a priority during preparations but turned out not to be problematic; perception of speed was rated neutral and its effect on performance marginal through all flights. However, touch-down speed was confirmed by three tail strikes as a problem; although not an initial concern it was commented during familiarization training. This subsequently prompted increased instructor attention on the risk of tail strike. Even though the ratings on the use of the side control suggest that it was not problematic from the first selected flight, the ratings improved in the subsequent flights. Regarding the work environment, the fact that dual instrumentation is not available on the aircraft seemed to have no negative effect on the training according to the instructors. Ratings on safety were neutral compared to the previous type of aircraft used at the school.

Improvement of training is the main expected benefit from the new aircraft, a training that more effectively produces competent first officers for large jet transport aircraft. The assumption seems to be that the smaller aircraft will be able to simulate the environment of a larger aircraft (use of PFD, increased focus on system management, use of side control and automation) and thus provide more relevant training for the student (Casner, 2003b). This assumption is questionable since it is not clear how different levels of fidelity in simulation connect to different levels of learning. Caird (1996) stated that, “for decades, the naïve but persistent theory of fidelity has guided the fit of simulation systems to training.” The potential transfer effect of initial pilot training with TAAs to later stages of pilot training needs to be investigated to see if it brings temporary or long term benefits to pilot’s skills.

Although the issue of transfer of training needs further attention, other benefits of using TAAs were more evident. The flight instructors were unanimously positive regarding the work environment in the aircraft and the effects they expected it to have on their own teaching as well as on the learning environment for the students. (Effects regarding less noise and vibrations and improved space were stressed, particularly when performing the fourth or fifth flight of the day.) The instructors were similarly convinced that training with modern displays would prepare the students better for their future work environment in a modern jet transport aircraft. Combined with the fact that the training progression was similar to that of previous courses and the limited problems with the use of the new technology, this indicates that TAAs can provide an improved learning environment combined with a potential for positive transfer of use of new technology to later stages of pilot training. However, the “doubled edged sword” of technology mentioned previously is present also in these benefits. While pilot and aircraft capabilities may be increased by the new displays and controls, the extent to which these may also provide a potential for increased workload and new ways for student pilots to fail is still not known.

While this study has shown that TAAs can be successfully introduced in ab-initio training, further research on the expected positive outcomes is required. The problem of trying to restrict use of the features available in a TAA was illustrated by the student engaging automation during one of the first solo flights. The transition to a traditional twin-engine aircraft in later stages of the training will also be of interest, since students trained on TAAs might begin their careers flying aircraft with traditional cockpits. Expressions of risk compensation should also be considered, where design features intended for increased protection and safety (parachute, moving map) is converted into mechanisms for accepting greater risk and smaller margins. In addition, the question of whether there are training qualities lost in the transition to TAAs should be investigated. The comments from the PPL examination on students being too focused on the instruments and lacking in technical knowledge and fuel planning can be interpreted as indications that information readily available on displays shortcut or truncate active information management; *the student knows how to work the system, but not how it works*. To ensure that the introduction of TAAs will bring the expected benefits to pilots' training and that important cognitive skills will not be lost, these issues should be addressed.

The main conclusion from this study is that with extensive preparation, introduction of TAAs in ab-initio training can be accomplished successfully, offering an improved learning environment as well as a potential for both increased safety and positive transfer of training with modern technology to later stages of pilot training. However, the confirmation of the expected benefits of increased safety and transfer of training as well as questions on what might be lost in the process need to be addressed by further research.

References

AOPA Air Safety Foundation (2005). Technically advanced aircraft - Safety and training. *AOPA Air Safety Foundation Special Report*. Frederick, MD: AOPA Air Safety Foundation.

Billings, C. E. (1997). *Aviation automation: The search for a human centered approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Caird, J.K. (1996). *Persistent issues in the application of virtual environment systems to training.* Paper presented at the Third Annual Symposium on Human Interaction with Complex Systems, 1996. Proc. HICS'96. Los Alamitos, CA: IEEE Computer Society Press, 124-132.

Casner, S.M. (2003a). Teaching cockpit automation in the classroom. *NASA report* NASA/TM-2003-211865. Moffett field, CA: Ames Research Center.

Casner, S.M. (2003b). Learning about cockpit automation: From piston trainer to jet transport. *NASA report* NASA/TM-2003-212260. Moffett field, Ca: Ames Research Center.

Craig, P.A., Bertrand, J. E, Dornan, W., Gosset, S., & Thorsby, K.K. (2005). Ab initio training in the glass cockpit era: New technology meets new pilots. *Proceedings of the 13th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Dekker, S.W.A., & Hollnagel, E. (1999). Computers in the cockpit: Practical problems cloaked as progress. In Dekker S. W. A. & Hollnagel E. (Eds.), *Coping with computers in the cockpit*, (pp. 1-6). Aldershot, UK: Ashgate.

Dekker, S.W.A., & Johansson, B. (2000). *JAR-FCL and pilot knowledge needs*. Report from the Swedish Centre for Human Factors in Aviation. Linkoping, Sweden: Swedish Centre for Human Factors in Aviation.

Dekker, S.W.A. (2004). *Ten questions about human error: A new view on human errors and systems safety.* Mawhah, N.J.: Lawrence Erlbaum Associates.

Federal Aviation Administration (2003). General aviation technically advanced aircraft FAA – Industry safety study. *Final report of TAA Safety Study Team*. Washington, DC: FAA.

Fanjoy, R.O., & Young, J.P. (2003). Advanced collegiate flight automation training: What is needed and at what cost? *International Journal of Applied Aviation Studies 3*(2), 215-225.

Fanjoy, R.O., & Young, J.P. (2004). Training levels and methodologies for glass cockpit training in collegiate aviation. *Journal of Aviation/ Aerospace Education & Research 13*(2), 35-41.

Fanjoy, R.O., & Young, J.P. (2005). Flight deck automaton: Line pilot insight for improved initial pilot training. *International Journal of Applied Aviation Studies 5*(1), 13-24.

Higdon, D. (2000). Plastic planes, part two: The Cirrus SR20. The Executive Summary: People Flock To Winners For A Reason. *AVweb,* November 13, 2000. Retrieved March 14, 2006 from http://www.avweb.com/news/newacft/182428-1.html

Nählinder, S., Berggren, P., & Persson, B. (2005). Increasing training efficiency using embedded pedagogical tools in a combat flight simulator. *Proceedings of the 49th annual meeting of the Human Factors and Ergonomics Society,* 2197-2200. Santa Monica, CA.

Rignér, J., & Dekker, S.W.A. (1999). Modern flight training – Managing automation or learning to fly? In Dekker S. W. A. & Hollnagel E. (Eds.), *Coping with computers in the cockpit*, (pp. 145-151). Aldershot, UK: Ashgate.

Roessingh, J.J.M., van Gent, R.N.H.W., van den Bosch, J.J., Biemans, M.C.M., van Bavelgem, J., Fletcher, G.C.L., Lodge, M., Koehl, F., & Linsenmayer, B. (1998). *European collaboration on transition training research for improved safety (ECOTTRIS) – Final report*. Report prepared for the European Commission DG VII (Transport) under contract N_: AI-96-SC.201. Report id: ECOTTRIS/NLR/WPR/WP41.0.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Utilizing Situational Judgment Tests (SJT) for Pilot Decision-Making

Brian G. Dillman

and

Jang R. Lee
Purdue University
Department of Aviation Technology
1 Purdue Airport
West Lafayette, IN 47907-3398

Abstract

Pilot decision-making is an important issue to flight safety. Students in collegiate flight training are pursuing careers as professional pilots, and what they initially learn about decision-making will guide their actions throughout their flying careers. Investigating aeronautical decision-making can be a tool for improving the decision-making quality of collegiate flight students and subsequent professional pilots. The more flight students are introduced to the process of aeronautical decision-making, the stronger their understanding of the limitations of their abilities. With the close link between flight safety and decision-making, it is only natural that the safety culture as a whole will be improved through this process. The purpose of this paper is not to determine the effectiveness or validity of the situational judgment test, but illustrate the utilization of the situational judgment test within a collegiate flight training program.

Introduction

For many individuals, both in aviation and other industries, safety is not a natural mentality. It is one that must be developed and learned through multiple and varying experiences. Likewise, decision-making in any field is a concept that is strengthened through experience and past occurrences. In aviation, the time period between a student pilot starting flight training and the point at which they

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.

reach a satisfactorily level of experience to draw upon when making decisions is a perilous period. Not surprisingly, in 2002, pilots with 500 or fewer total hours accounted for 37.3% of all accidents, and 30.8% of fatal accidents (Hummel, Murphy, & Wright, 2003). That number increases to 53.0% of all accidents and 47.5% of fatal accidents for pilots with less than 1000 hours (Hummel, Murphy, & Wright, 2003). O'Hare (2003) said that decision-making is the key issue to the question of effective human performance in aviation. Jensen (1997) said the term "aeronautical decision-making" (ADM) has been used to describe and assess pilot judgment within many aviation circles. In addition, Jensen mentioned that "Unfortunately, most pilots do not receive structured decision or judgment training either in their initial or later flying experiences" (p. iv). It is assumed that they will learn judgment through their experience (Inagaki, Takae, & Moray, 1999).

There needs to be a way to show or illustrate to these pilots that they do not "know it all", and that there is still significant room for improvement. By coming to grips with an individual's deficiencies, they will cautiously respond when dealing with situations that may arise during flight training or during the flight time building phase of their careers. Aviation educators are interested in how flight students make decisions during critical situation to flight safety, and how close their decisions are to those of industry accepted experts. With the knowledge of deficiencies in decision-making skills, safety officers and other individuals responsible for flight safety can develop tools to encourage and strengthen areas of deficiency.

Decision Making

Decision-making is the choosing of issues that require attention, setting goals, finding or designing suitable courses of action, and evaluating or choosing among alternative actions (Simon, 1983).

Pilot decision-making has been regarded as one of the most important processes affecting safe flight operations and also called as aeronautical decision-making (Jensen, 1995). Hunter (2003) stated that the terms of aeronautical decision making (ADM) and judgment may be used interchangeably as they are similar in scope and meaning.

Dreyfus and Dreyfus (1986) stated that the primary difference between competent performers and experts in aviation is their decision making/judgment. In their model, there is deficient judgment in novice and advanced beginner stages. The competent performer judges by means of conscious deliberation. Proficient and expert performers make judgments based on prior experiences in ways that cannot be explained. It is this lack of prior experiences that flight students with low flight time struggle with to make correct decisions, but in which they can also learn from the process.

Jensen (1995) felt that judgment is learned, primarily, from experience and with lack of experience is it difficult, if not impossible, to learn judgement. "In reality, the best way to learn judgment is to discover it from the experiences of others, and although experience is a great teacher, in aviation, the experience of others is safer" (Jensen, 1995, p.176).

Statement of the Problem

Collegiate flight training deals with students at the average age range of 18-21 years old. There are many traits commonly used to describe this age group; invulnerable, naïve, impulsive, energetic, eager, and inexperienced to name a few. It is traits like these, along with historical references, that have prompted automobile insurance companies to charge higher premiums for this age range. This same phenomenon needs to be seriously considered when dealing with collegiate flight students who have those same traits plus a low level of flight experience. This reduced level of experience leads to a lack of decision-making skills since they have very little past experience to reflect upon when making decisions. It is our job, as professional educators, to introduce flight students to as many different scenarios as possible so that they can utilize those experiences later in life when similar situations arise. Unfortunately, it is very expensive to experience these scenarios in an airplane, and that cost is only marginally reduced by the use of simulators or training devices. Ideally there should be a low-cost tool that would introduce the novice pilot to scenarios that have occurred in the past, define possible courses of action, and then let the pilot use their decision making skills to select the best possible choice. Just such a tool has been developed and implemented by David Hunter of the Federal Aviation Administration. Hunter (2003) reported that the Situational Judgment Test (SJT) has the potential for use in the assessment of judgment or aeronautical decision-making by general aviation pilots. With the situational judgment test, the opinions and experience of industry experts can be a fantastic gauge to whether students, with their limited experience, are moving in the right direction when it comes to decision-making skills. SJT consists of fifty-one scenario-based questions asking how the pilot would make a decision. Each question has four alternative solutions and each solution is ranked from first to fourth by a panel of subject matter experts (SMEs). Hunter completed two separate studies utilizing a situational judgment test to assess decision-making skills. The first study was completed with a paper-based format and the second was completed utilizing an internet-based format. In both studies the completion requirement was voluntary and the median age was 47 for the first study and 45 for the second with a standard deviation of 13 for both studies (Hunter, 2003). As the age of an individual increases, the perceived need for personal development and improvement increases, and this can be seen from the response rates that were obtained in the two studies from the potential age groups that were eligible for completion of the situational judgment test. Unfortunately in collegiate aviation the average age is significantly lower than the test sample with which Hunter worked, and it was the intent to make this decision-making analysis mandatory rather than voluntary. Due to these changes, it was unclear whether the situational judgment test would prove beneficial in this circumstance.

Method

Flight students in the Purdue University's Professional Flight Curriculum working on their Instrument Ratings and Private or Commercial Certificates are required to attend bi-weekly meetings for the purpose of discussing safety related topics on a continual basis throughout the given semester. In September of 2004 the Situational Judgment Test was disseminated at the safety meeting to all of the flight students in attendance. The intent of disseminating the Situational Judgment Test to the students in the flight curriculum was to allow them to think about various scenarios and use their current level of experience to come up with a response.

Within Appendix A there are five questions from the fifty-one total that were disseminated to the flight students during the initial testing phase. The ranking by the industry experts for each possible courses of action is within parenthesis next to the letter choice. The number one indicates the most favorable choice by the industry experts and the number four indicates the least favorable choice. As can be seen, the questions have a significant amount of reference information and that material must be processed for the flight student to formulate an appropriate action based upon their previous experience. At the completion of the initial testing, the scores for each student were tabulated based upon the opinions of the industry experts.

Delivery Adjustments

Two main issues were encountered during the initial delivery that were not expected. The first issue was trying to convey the fact that this assessment was to make them better pilots and that they were not actually required to complete the assignment. College students are very adept at determining whether an assignment will have any bearing on grades or course success. “It will make you a better person” or “It will make you a better pilot” is sometimes not a good enough reason for some students to dedicate enough time to the successful completion of the assessment device. At the beginning of the assessment, the students believed that there would be a grade for the completion of the task. At various points the students determined that there would be no grade for the test and either quit filling it out completely or randomly started to select answers so that they could “complete” the assignment. Some of the students placed a higher value on leaving the meeting to go home or get to another scheduled meeting than successfully answering each scenario with the best possible choice.

The biggest issue with the implementation of the judgment test was that there is no right or wrong answer. Each scenario has several possible responsive actions that are ranked from the most appropriate to the least appropriate choice. In most university courses a properly written assessment tool has a right answer and one or more incorrect answers. If someone knows the correct answer to a question then it is apparent that the other choices are incorrect. If someone does not know the correct answer, then all choices, if the question is written properly, will appear plausible and correct. In this assessment tool there was no clear correct answer so several students struggled with determining what answer to select as their responses. It was difficult for some students to grasp the idea that they should select the answer that most closely matched how they would respond if placed in the exact same scenario.

Due to the unforeseen problems with the first dissemination of the SJT, it was decided that the test would be delivered a second time with a little more explanation and preparation to eliminate misunderstandings. Although the students were not thrilled at the prospect of spending another hour of time completing the assessment tool, they were more willing to be active participants due to the explanation of the significant gains in aeronautical knowledge can be achieved by the test administrators. At the completion of the second testing, the scores for each student were tabulated and compared to the initial testing. The scores for both testing periods, along with the amount and percent increase or decrease, are listed in Appendix B.

Results

If the first delivery of the assessment tool is considered as a pretest and the second delivery as a posttest, it can be seen that there was an improvement of the overall scores. Scores went from an average percentage score of 78% to an average of 82%, which is an increase of 4%.

Many of the participants were interested to see how they scored compared with the industry experts. In many cases the opinions of the experts served as reinforcement to the students' thought processes and other times it was a wake-up call for remedial training. The most interesting dialogue occurred when students saw the rankings by the experts for each possible course of action and wanted further clarification as to why the experts ranked the choices in a certain order. The flight department was not privy to the thought process behind the rankings by the experts so this was a perfect opportunity to force the students to use critical thinking skills to rationalize why the experts ranked the items in the given order. In many cases this proved to be a perfect opportunity for guided discovery learning to occur which was beneficial for the students.

Discussion

Did the introduction of the Situational Judgment Test improve or enhance decision-making within the flight training department at Purdue University? It was not the intent of this project to determine whether the SJT that was utilized had a statistically significant impact in the decision-making skills of the participants, but to illustrate one tool for utilization in increasing student pilot experience without placing individuals in danger. By utilizing this device it prompted the students to at least consider the deficiencies they possess in this area of piloting skill. Giving a test on decision-making in flight scenarios to flight students forces them to consider situations that they have never encountered. It is probable that they had very little on which to base their decisions, but that is the process that builds their experiences. It can be seen that the average overall score increased on the second delivery of the SJT. This increase could be due to a number of reasons, such as improved directions for taking the test or better understanding of the expectations by the students. It is also possible that the students took the first test, reflected on the possible choices for several of the scenarios that they had never considered, talked with friends or instructors, and then utilized that understanding to answer the question on the second test.

The more flight students are introduced to the process of aeronautical decision-making the stronger their understanding of the limitations of their abilities. With the close link between flight safety and decision-making, it is only natural that the safety culture as a whole will be improved through this process. There are ways to provide the training required to produce a pilot by Federal Aviation Regulation standards. It is devices, such as the situational judgment test, that can assist in the development of pilots that utilize critical thinking to enhance their decision-making processes.

Appendix A

* Cessna 172 (1) Data

Type: Four-seat light aircraft
Engine: One flat four piston engine of 160 hp

DIMENSIONS:
- Wing span: 35 ft 10 in
- Length: 26 ft 11 in
- Height: 8 ft 10 in

WEIGHTS:
- Empty: 1,430 lb
- Max. takeoff: 2,300 lb

PERFORMANCE:
- Max. speed: 125 kt
- Max. cruise: 122 kt
- Initial climb: 770 ft per min.
- Service ceiling: 14,200 ft
- Max. range: 575 nm with 45min reserve

SURVEY QUESTIONS

1. You are flying an "Angel Flight" with a nurse and non-critical child patient to meet an ambulance at downtown regional airport. You filed VFR, it is 11:00 P.M. on a clear night when at 60 NM out you notice the ammeter indicating a battery discharge and correctly deduce the alternator has failed. Your best guess is that you have from 15 to 30 minutes of battery power remaining. You decide to:

   A (4). Declare an emergency, turn off all electrical systems except for 1 NAVCOM and transponder and continue to the Regional Airport as planned.

   B (1). Declare an emergency and divert to the Planter's County Airport which is clearly visible at 2 o'clock, 7 NM.

   C (3). Declare an emergency, turn off all electrical systems except for 1 NAVCOM, instrument panel lights, intercom and transponder and divert to the Southside Business Airport which is 40 NM straight ahead.

   D (2). Declare an emergency, turn off all electrical systems except for 1 NAVCOM, instrument panel lights, intercom and transponder and divert to Draper Air Force Base which is 10 o'clock at 32 NM.

Airport	Runway	24hr Tower	Class C	Lightened R/W	Telephone Available	Maintenance
Regional Airport	8800x150 7753x150	Yes	Yes	Yes	Yes	24 hrs
Planters County Airport	3200x75	No	No	Yes	Yes	0700-1800
Southside Business Airport	4835x100 4129x100	Yes	Yes	Yes	Yes	0700-1800
Draper AFB	11500x300	Yes	No	Yes	Yes	None

2. You are solo on a late night cross country cruising VFR at 9500 feet with two hours left to your destination when you become very drowsy. You decide to:

A (4). Direct the cold air vent onto your face, sing, keep moving about, anything you can to keep awake.
B (1). Land at an airfield 8 miles ahead, get a motel room and call it a night.
C (3). Descend and continue flying at a lower altitude.
D (2). Land at the airstrip ahead, walk around, then takeoff and continue.

3. In the evening after an exhausting three day business meeting at a downtown hotel, you have loaded your rental airplane at the Downtown Airport and prepare to file your VFR flight plan for the two hour flight home when you discover you left your only pair of reading glasses in the meeting room back at the hotel. You have no problem seeing the panel gages, or distance vision, but can't read a map or chart. Weather is solid VFR and if you depart within the next 20 minutes you will arrive at your home airport before dark. You decide to:

A (4). Depart and fly home.
B (2). Call the hotel, if they have your glasses, go get them and fly home late this evening.
C (3). Call the hotel, if they do not have your glasses, spend the night, have a pair expressed to you and fly home tomorrow.
D (1). Call the hotel, if they have your glasses, go get them, spend the night and fly home in the morning.

Airport	Runway	24hr Tower	ARSA	Lightened R/W	Telephone Available	Maintenance
Regional Airport	8800x150 7753x150	Yes	Yes	Yes	Yes	24 hrs

4. You are cruising at 4500 feet on top of a thin haze layer with the outside air temperature at 65 degrees. It has been twenty-five hours since the engine was overhauled and the run-up check was well within limits. The engine slowly loses RPM with no indications of oil or fuel problems. You suspect carburetor icing and pull on the carb heat. The engine backfires, vibrates and loses RPM fast. You decide to:
   A (4). Pull out the mixture, stop the engine and check the fuel selector valve, mag switch settings and declare an emergency.
   B (2). Push in the carb heat, keep the engine running and divert to the closest airfield.
   C (1). Keep the carb heat on and see what happens.
   D (3). Push in the carb heat, keep the engine at idle, declare an emergency and ask for advice.

5. You are preparing to enter the VFR traffic pattern at the Regional Airport and hear the tower report winds from 280 at 15 knots, and they are vectoring traffic to the primary 8800 ft runways 35. A Piper Cherokee asks to use the 7753 x 150 runway 27. The Cherokee is told the runway is not active, but to you it looks OK. You decide to:
   A (4). Accept clearance to runway 35 and follow the traffic.
   B (2). Ask to use runway 27.
   C (3). Insist on using runway 27 stating that the crosswinds are unsafe for you to use runway 35.
   D (1). Divert to the Southside Business Airport where the runway is almost directly aligned with the wind.

Airport	Runway	24hr Tower	Class C	Lightened R/W	Telephone Available	Maintenance
Regional Airport	8800x150 7753x150	Yes	Yes	Yes	Yes	24 hrs
Southside Business Airport	4835x100	Yes	Yes	Yes	Yes	0700-1800

Appendix B

Student No.	1st SJT Raw Score	2nd SJT Raw Score	Difference (Raw Score)
1	168	176	8
2	175	180	5
3	156	175	19
4	161	176	15
5	178	181	3
6	154	162	8
7	167	176	9
8	146	171	25
9	161	174	13

10	166	183	17
11	145	166	21
12	150	164	14
13	158	162	4
14	178	174	-4
15	173	172	-1
16	171	166	-5
17	152	171	19
18	162	177	15
19	171	173	2
20	176	181	5
Overall (For All Students)	158.85	167.95	9.49

Student No.	**1st SJT (% Score)**	**2nd SJT (% Score)**	**Difference (%)**
1	0.82	0.86	0.04
2	0.86	0.88	0.02
3	0.76	0.86	0.09
4	0.79	0.86	0.07
5	0.87	0.89	0.01
6	0.75	0.79	0.04
7	0.82	0.86	0.04
8	0.72	0.84	0.12
9	0.79	0.85	0.06
10	0.81	0.90	0.08
11	0.71	0.81	0.10
12	0.74	0.80	0.07
13	0.77	0.79	0.02
14	0.87	0.85	-0.02
15	0.85	0.84	0.00
16	0.84	0.81	-0.02
17	0.75	0.84	0.09
18	0.79	0.87	0.07
19	0.84	0.85	0.01
20	0.86	0.89	0.02
Overall (For All Students)	0.78	0.82	0.04

References

Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* New York: The Free Press.

Hummel, K., Murphy, K., & Wright, D. (2003). 2003 Nall Report: Accident Trends and Factors for 2002. Frederick, MD: Aircraft Owners and Pilots Association: Air Safety Foundation. Retrieved April 1, 2005, from http://www.aopa.org/asf/publications/03nall.pdf.

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

Inagaki, T., Takae, Y., & Moray, N. (1999). Automation and human-interface for takeoff safety. *Ninth International Symposium on Aviation Psychology*.

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

Jensen, R. S. (1997). A treatise on the boundaries of aviation psychology, human factors, ADM, and CRM. In R. S. Jensen (Ed.). *Ninth International Symposium on Aviation Psychology (pp. iv – vii).* Columbus, OH: The Ohio State University.

O'Hare, D. (2003). Aeronautical decision making: Metaphors, models, and methods. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 201-237). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Simon, H, A. (1983). *Reason in human affairs.* London: Basil Blackwell.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Computer-Based Flight Simulation: A Cost Effective Way for General Aviation Pilots to Improve Their Instrument Proficiency

Jon McDermott
Bowling Green State University
College of Technology
Bowling Green, OH 43403
Phone: 419-372-7563
Email: jmcderm@bgsu.edu

Abstract

Aviation educators have long sought economical and effective ways to simulate how an aircraft flies. In particular, there has always been an interest in simulating the procedures and practices of operating an aircraft in instrument flight conditions, when no horizon is visible. As the price of flying aircraft increase, there has been an interest in adopting the technology of the microcomputer to bring realistic flight simulation at lower costs to the general aviation market. This paper offers the results of a study of pilot instrument proficiency and how computer-based flight simulation can improve pilot performance. Although practicing instrument approach procedures in a simulation device is not a requirement for recent flight experience under Federal Aviation Regulation 61.57, this study proves that such practice is a valuable resource that pilots need to consider when maintaining, and improving, their instrument flight skills.

Introduction

Aviation educators have long sought economical ways to duplicate or simulate the human processes necessary to operate safely an aircraft in flight and on the ground in a device other than an actual aircraft. The reasons for providing such simulations are many, but in particular represent a cost effective way to train and educate pilots as the costs associated with operating aircraft increase. The capability to simulate actual flight conditions began in 1934 with the C-3 Link Trainer (or "blue canoe" as it was called by the military pilots who used it) and grew in popularity through the 1980s. This has resulted in the construction of

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.

multimillion dollar, multifunctional simulation marvels that remain a flight training foundation for every branch of the military as well as the major airlines worldwide (Williams, 1994).

However, recent advances in software and hardware systems have allowed engineers to build much smaller, more cost effective computer devices that provide as accurate a simulation of flight conditions as those bulky, stand-alone facilities of the 1980s. In 1989, it was recognized that "within the next few years, low cost computers . . . will be used to supplant most, if not all, training for . . . private, commercial and instrument-rated pilots as well as the various air traffic control positions" (Connolly & Lehrer, 1989, p. 443).

Expectedly, there has been an increase in the academic examination of low cost flight simulation devices (Johnson & Stewart, 2002; Taylor et al., 2001; Taylor et al., 2003; Taylor, Talleur, Rantanen, & Emanuel, 2004). As computer software capabilities improve, and hardware configurations better represent the flight deck environment of actual aircraft, computer-based simulation devices have become a valuable tool for presenting realistic, high-quality representations of aircraft performance and instrumentation.

Thomson, Lintern, and Brady (as cited in Taylor et al., 2003) documented the benefits of simulating aviation operating practices in simulation devices and the transferability of learned skills to actual aircraft operations. Carretta and Dunlap (1998) in their Air Force Research Lab study found that pilot skills learned in a simulator were easily transferred to the actual aircraft. These studies attest to the validity of simulation as a valid format for pilots to practice their flight skills.

Simulation as Computer Games

The introduction of computer gaming that simulates the actual flight conditions a pilot can experience is a natural extension of computer game playing for educational purposes. The military, which began simulation, albeit in a rudimentary fashion with the "blue canoe" as earlier referenced, quickly adapted the efficiency, power, and flexibility of modern computer software and hardware for use in its flight education programs. Subsequently, what began as a military adventure into simulating the high performance and sophistication of the actual tactical operations of its highly sophisticated aircraft was quickly adopted by commercial airline education programs as this industry began to operate its own versions of high performance jet aircraft.

Shortly thereafter, general aviation, the population that operates less sophisticated propeller aircraft, began pursuing simulation devices based on gaming principles that became the foundation upon which novice aviation enthusiasts could practice pilot flight skills. Roscoe (1971) studied the degree to which a task learned in a simulator could be interpolated to other learning, to an aircraft for instance, specifically when comparing the rate of transfer of learning to a control group that had no previous training. He found that the transfer effectiveness of simulation differs for the types of training accomplished; simulation that may be ineffective for pre-solo training is effective for instrument and cross-country training. He added that the most useful measure for the educator in determining the amount of simulation needed is when the cost of such training saves the expense of more costly activities such as actually flying an aircraft. Synthetic learning devices (simulators) have always offered general aviation an economical way to improve the safety and effectiveness of aviation training activities (Lintern, Roscoe, & Sivier, 1990). In particular, Lintern et al. (1990) studied the use of simulators to transfer visual maneuvers, finding "skills learned in the simulator are relevant to real flight and that instructional procedures . . . are likely to . . . transfer to the control of an airplane" (p. 302).

Effectiveness of Simulation

Hampton (1991) studied personal computer (PC) based simulation effectiveness in training private pilot candidates. He cited a study by Caro in 1972 that showed simulators were effective in transferring trained skills to the aircraft. He also cited Gerhard's 1983 study that determined that simulators reduce training costs and provide an inherently safe learning environment for pilot training. Hampton's own study indicated that PC-based simulation was as effective as flight training devices (FTDs), $F(2, 27) = 3.27$, $p < .05$.

Embry Riddle Aeronautical University also investigated the use of personal computer simulation devices to reduce the costs of flight education. Specifically, the university developed AGATE, or the Advanced General Aviation Transport Experiment (Collins, 2000). The first group of AGATE students, utilizing desktop computer devices to supplement both visual and instrument training, "reduced flight time needed by 29% and the costs of training by 20%" (Kocks, 1998, ¶ 20). The National Aeronautics and Space Administration followed the AGATE experience by developing its own integrated flight training process, the Small Aircraft Transportation System (SATS), which continued to utilize desktop computer devices in its integrated private and instrument training process (Collins, 2000).

Allerton's (2000) study cited Taylor's research into 33 studies between 1939 and 1977 indicating that there was a positive transfer of training and that simulation improved the quality of training. Lintern (1992) studied the rate at which skills were transferred in simulation and learning games. He cited Thorndike's 1903 claim that transfer of learning occurs when two common tasks are related in general terms, noting that "almost 90 years later we are still struggling to characterize and to identify the elements that support transfer" (p. 338). He cited his own work in 1991 that proves that low fidelity simulation is as effective in transferring flight skills as high fidelity.

Carretta and Dunlap (1998) investigated the training effectiveness of simulators from 1986 to 1997. They cited Hay, Jocobs, Prince, and Salas' 1992 work that reviewed 247 journal articles, book chapters, and technical reports regarding training effectiveness. This meta-analysis revealed several trends that were not readily apparent from many studies, including that simulators consistently led to improved training effectiveness in jet pilots. They cited recent studies by Lintern and Garrison (1992), Lintern, Roscoe, Koone, and Segal (1990); Lintern, Roscoe, and Sivier (1990); Lintern, Sheppard, Parker, Yates, Nolan, and Roscoe (1987); Pfeiffer, Horey, and Butrimas (1991), Taylor, Lintern, and Koonce (1993); Weastra et al. (1986), and Wightman and Sistrunk (1987). These studies showed simulators were useful in training various flight skills, with increasing effectiveness as the amount of simulator training increased. Carretta and Dunlap added that recent developments in computer technology have led to significant improvements in simulator computerization with faster, more powerful systems operating appropriate simulator hardware.

Finally, Johnson and Stewart (2002) of the U.S. Army Research Institute investigated the effectiveness of a PC-based simulation device at improving helicopter training processes. Their study of 16 aviators utilizing a utility method of inquiry found that the "micro-computer was valuable in supporting the training of navigation instruments and procedures" (p. 13); nonparametric: Sign test $N = 6$, $\chi = 0$, $p < .02$ for experienced aviators; nonparametric: Sign test, $N = 10$, $\chi = 0$, $p < .001$ for student aviators.

Participants

The participants of this study included general aviation pilots who held at least a Federal Aviation Administration (FAA) Private Pilot Certificate with an instrument rating and who were randomly selected from an FAA database of pilots in the northwest Ohio region. Although the selection process was primarily based on convenience, the researcher utilized suggestions by Watters and Biernacki (1989) to target this population. According to Watters and Biernacki, target sampling is a "purposeful, systematic method by which a controlled list of specific populations within a geographic area are recruited" (p. 420). They are not random samples but reflect "a strategy to obtain systematic information when true random sampling is not feasible" (p. 420). O'Connell (2000) added that a targeted sample selects participants with "specific attributes important to the subject under study" (p. 223), in this study pilots with instrument ratings. She concluded that such a sample represents "a useful methodology for constructing replicable samples. . . that maintains a strong congruence to the targeted population" (p. 224).

To increase the validity of this nonprobable sample, the researcher also utilized random number generating software to identify the target sample from the FAA database. According to O'Connell (2000), the use of randomization can strengthen the validity of study results. Randomization also tends to "balance out the effects of extraneous factors evenly across groups and offers protections against threats to internal validity" (p. 228).

Apparatus

One FAA approved Frasca FTD Model 141, serial number 129, Urbana, Illinois, and one Precision Flight Controls Personal Computer Aircraft Training Device Model PI 142, serial number 13831, Mather, California, were used as simulator training devices in this experiment.

Design

This study represents an experimental examination of pilot proficiency. Thirty-four participants were randomly selected from the list provided by the FAA as identified earlier. This was a convenience sample, supplemental with a randomized targeting type of strategy to emphasize consistency between the sample and the population while attempting to account for bias as much as possible (O'Connell, 2000). Participants each completed a pilot questionnaire indicating their flight and simulation experience. The purpose of the pilot questionnaire was to assist the researcher in determining the experience level of the pilot participants and to what degree, if any, was their simulation experience.

In this study, the randomly assigned group of participants accomplished a 15-minute practice session of basic instrument flight skills to familiarize themselves with the simulation device. Each participant then accomplished, in a series, four instrument landing system approaches with a procedure turn, with proficiency on the first and last simulated approach assessed on a 4-point scale ranging from the pilot being unable to perform the maneuver to the pilot performing at a skill level where no deviations took place.

A statistical analysis was performed on the outcomes of each assessment to ascertain if simulation experience had a positive effect on participant instrument proficiency as demonstrated on the simulation devices used in this study. A second analysis was performed to determine if participant instrument proficiency improved as a result of accomplishing, or practicing, the series of four approaches with a procedure turn. After the simulations, each participant was asked to provide feedback on this simulation experience and whether they thought their instrument proficiency improved.

Results

The pool of participants was very diverse, ranging from professional airline pilots who had hundreds of hours of multiengine experience in aircraft and simulation devices to single engine aircraft, general aviation pilots who had never flown a simulation device before. Table 1 identifies pertinent demographic information concerning the participant pool. Nearly one fourth of the participants had logged over 1,500 hours of total flight time, a significant level of flight experience equivalent to certification as an Airline Transport Pilot (*Federal Aviation Regulations and Aeronautical Information*, 2005, p. 107). However, almost one third of the participant pool had less then 10 hours of actual instrument flight experience. Almost one-half of the pool had less then 10 hours of simulator experience.

Table 1
Demographics of Participants

Flight Experience of Participants	% of Participants
Total flight experience <250 hours	17
Total flight experience >1500 hours	24
Actual instrument flight experience <10 hours	32
Actual instrument flight experience >250 hours	16
Simulator experience <10 hours	41
Simulator experience >250 hours	11
Flight experience (last year) <10 hours	14
Flight experience (last year) >250 hours	17
Instrument approaches (last 6 months) <3	24
Instrument approaches (last 6 months) >12	47

Table 2 identifies the results of an analysis of variance (ANOVA) of the pilot's performance on the first assessment versus the pilot's simulation experience. It indicates there was a significant difference between pilot instrument proficiency of those who had simulation experience and those who did not, $F(1, 31) = 4.5$, $p < .05$.

Table 2
Analysis of Simulation Experience and First Assessment

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	819.621	2	409.811	4.046	.027
Within Groups	3140.143	31	101.295		
Total	3959.765	33			

Table 3 identifies the results of an ANOVA of the pilot's performance on the second assessment (fourth simulation experience) versus the pilot's simulation experience. Again the data indicates that there was a significant difference between instrument proficiency of those pilot's who had simulation experience and those who did not, $F(1, 31) = 3.6$, $p < .05$.

Table 3
Analysis of Simulation Experience and Second Assessment

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	534.845	2	267.423	3.645	.038
Within Groups	2274.213	31	73.362		
Total	2809.059	33			

Table 4 identifies the results of an ANOVA of the difference between, or improvement in, simulation assessment scores between pilots with and pilots without simulation experience. The data indicates no significant difference between the rate of improvement, or lack thereof, of participants with simulation experience and those without. There was no significant difference in the rate of improvement in instrument proficiency after the participant had accomplished four approaches with procedure turns, $F(1, 31) = 1.2$, $p > .05$.

Table 4
Analysis of Difference Between First Assessment and Second

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	35.800	2	17.900	1.201	.314
Within Groups	461.965	31	14.902		
Total	497.765	33			

Summary

Lintern et al. (1990) reported that flight skills learned in a simulator are relevant to real flight in an aircraft and therefore are transferable to the aircraft. Participants agreed with Lintern, Roscoe, and Sivier. In particular, one participant commented that even though the simulators were more sensitive than his aircraft, each approach he flew improved and will make his aircraft easier to fly, agreeing with the transferability of this type of learning. Participants also appear to agree with Lintern et al. that learning took place in the simulator devices. Table 5 depicts the percentage of participants who thought that their pilot instrument flight skills improved while practicing ILS approaches during this study. Eighty-six percent of participants thought that their ability to fly an ILS approach improved. This improvement was closely followed by 84% of participants believing that their interception of a localizer course, a critical flight skill for accomplishing an ILS, improved. Other instrument skills that participants believed improved were navigation tracking, intercepting a glideslope, and accomplishing a procedure turn. Accomplishing a missed approach showed the lowest level of improvement, but still over one-half of the participants in this study thought that their instrument flight skills improved on this task. Concurring with Lintern et al., participants responded that their flight skills improved and would be transferable to aircraft operations.

Table 5
Participants' Feedback on Skill Improvement

Task	%
Airspeed	59
Altitude	78
Attitude	73
Nav Track	83
Procedure turn	76
Intercept localizer	84
Intercept glideslope	78
Accomplish ILS	86
Missed approach	63

Roscoe (1971) offered that learned tasks in a simulator can be transferred to an actual aircraft and that such a transfer follows a logical path—that of tasks learned and practiced in the simulator become learned and practiced operations in an actual aircraft. Participant comments agree with Roscoe's assertion that practicing instrument flight skills, in the case of this study in a simulation device, will improve their performance in an actual aircraft. One participant also commented that simulations are important in keeping an instrument pilot proficient without having to actually fly an aircraft. Feedback provided by participants indicated that 75% of participants *strongly agreed* that simulation is important in retaining instrument proficiency; another 21% *agreed* with that statement. This feedback not only attests to the cost effectiveness of computer simulation, but that such simulation devices may be appropriate as an alternative to flight experience in "actual or simulated instrument conditions, either in flight . . . or in a flight simulator or flight training device that is representative of the aircraft . . . for the instrument privileges sought." (*Federal Aviation Administration*, 2005, p. 59). Additionally, this feedback addresses the safety of general aviation pilot activities, an additional topic for FAA consideration.

Data from this study indicate that the general aviation instrument rated pilots who had simulation experience were significantly more proficient in accomplishing approaches with procedure turns than pilots without any type of simulator experience. The data also indicate that there was no significant difference in the rate of improvement in pilot instrument proficiency after practicing instrument approaches in either participant group, indicating that both pilots with and pilots without simulation experience improve equally when practicing a series of instrument procedures.

Considering the reality that the costs of flying aircraft are ever increasing, utilizing computer-based simulation devices to maintain and, when needed, improve a pilot's instrument proficiency is a worthwhile investment of a pilot's time and money. Although such flight experience in these simulation devices may not meet the requirements of federal regulations, this study proves that pilots can maintain, and even improve, their flight skills utilizing computer-based simulators, a benefit that may well save a pilot's life some day.

References

Allerton, D. J. (2000). Flight simulation—Past, present, and future. *The Aeronautical Journal, 104*(1042), 651-663.

Carretta, T. R., & Dunlap, R. D. (1998). *Transfer of training effectiveness in flight simulation: 1986 to 1997.* Brooks Air Force Base, TX: Air Force Research Laboratory, Human Effectiveness Directorate.

Collins, M. P. (2000). Tomorrow's training: Surprises are in store for flight instruction. *AOPA Pilot Magazine 43*(4). Retrieved January 31, 2005, from http://www.aopa.org/pilot/features/future0004.html

Connolly, T. J., & Lehrer, H. R. (1989). Applications of state-of-the-art educational technologies to prepare tomorrow's aviation professionals. *Proceedings of the 1989 Conference on Technology and Innovations in Training and Education* (pp. 443-448). Atlanta, GA: Embry-Riddle University.

Federal Aviation Administration. (2005). *Federal Aviation Regulations and Aeronautical Information Manuel.* New Castle, WA: ASA.

Hampton, S. (1991). PC-based simulation: An alternative for teaching instrument flying skills. *Technical Series, American Society of Manufacturing, 4,* 13-17.

Johnson, D. M., & Stewart, J. E. (2002). *Utility of a personal computer aviation training device for flight training* (Research Report 1787). Fort Rucker, AL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kocks, K. (1998). *Filling the 'qualification gap.'* Retrieved January 31, 2005, from http://www.defensedaily.com/cgi/rw/show_mag.cgi?pub+rw&mon=1298&file=overview1298.htm

Lintern, G. (1992). Flight simulation for the study of skill transfer. *Simulation/Games for Learning, 22,* 336-350.

Lintern, G., Roscoe, S. N., & Sivier, J. E. (1990). Display principles, control dynamics, and environmental factors in pilot training and transfer. *Human Factors, 32*(3), 299-317.

O'Connell, A. A. (2000). Sampling for evaluation. *Evaluation & the Health Professionals, 23,* 212-234.

Roscoe, S. N. (1971). Incremental transfer of effectiveness. *Human Factors, 13,* 561-567.

Taylor, H. L., Talleur, D. A., Bradshaw, G. L., Emanuel, T. W., Jr., Rantanen, E., Hulin, C. L., et al. (2003). *Effectiveness of personal computers to meet recency of experience requirements* (DOT/FAA/AM-03/3). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Taylor, H. L., Talleur, D. A., Emanuel, T. W., Jr., Rantanen, E., Bradshaw, G. L., & Phillips, S. I. (2001). Incremental training effectiveness of personal computers used for instrument training. *Proceedings of the Eleventh International Symposium of Aviation Psychology* (pp. 1256-1260). Columbus: The Ohio State University.

Taylor, H. L., Talleur, D. A., Rantanen, E., & Emanuel, T. W., Jr. (2004). *The effectiveness of a personal computer aviation training device, a flight training device, and an airplane in conducting instrument proficiency checks* (Final Report AHFE-04-12/FAA-04-5). Oklahoma City, OK: Federal Aviation Administration, Civil Aeromedical Institute.

Watters, J. K., & Biernacki, P. (1989). Targeted sampling: Options for the study of hidden populations. *Social Problems, 36,* 416-427.

Williams, K. W. (1994). *Summary proceedings of the joint FAA-Industry Conference on the development and use of PC-based aviation training devices* (DOT/FAA/AM-94/25). Oklahoma City, OK: Federal Aviation Administration, Civil Aeromedical Institute.

International Journal of Applied Aviation Studies, Volume 6, Number 1

The Effects of Hazardous Attitudes on Crew Resource Management Skills

Michael Wetmore

and

Chien-tsung Lu
Department of Aviation
T.R. Gaines 210E
Central Missouri State University
Warrensburg, MO 64903
Tel: 660-543-4458
Email: mwetmore@cmsu1.cmsu.edu
Email: ctlu@cmsu1.cmsu.edu

Abstract

The purpose of this investigation was to examine the relationship between aeronautical decision-making (ADM) and crew resource management (CRM) skills in general aviation accidents that involved a fatality. Factual Reports from the National Transportation Safety Board aviation accident database were analyzed in detail for evidence of pilot behaviors reflective of hazardous attitudes and pilot actions indicative of poor CRM skills. This study found relatively high frequencies of occurrence for certain ADM and CRM factors: (a) risk-taking; (b) hazardous attitudes; (c) poor decision-making; (d) pilot error; and (e) failure to utilize all available cockpit resources. One-way analysis of variance statistical tests were used to reject the null hypothesis that accident pilots with hazardous attitudes are as proficient at ADM and CRM skills as accident pilots without hazardous attitudes. This research paper concluded that hazardous attitudes have a measurable, negative effect on a pilot's ADM and CRM skills that can be summarized as follows: (a) more willing to accept high-risk flights; (b) more prone to making bad decisions; (c) more likely to commit pilot errors; and (d) less likely to use all of the available cockpit resources. We propose that research such as this can be used to create effective scenario-based and case-based teaching protocols in the aviation education and flight training environments.

Introduction

The purpose of this investigation was to study the relationships between pilot behaviors and pilot actions. More specifically, this paper examined the relationship between certain aeronautical decision-making (ADM) and crew resource

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.

management (CRM) factors in general aviation accidents that involved a fatality. Our goal was to provide aviation educators, flight instructors and student pilots a better understanding of the link between pilot behaviors and pilot actions. We hope that an increased level of awareness of how pilot behavior relates to pilot actions can help all of us in the aviation community to become safer pilots.

Data for this study was generated by a careful analysis of the National Transportation Safety Board (NTSB) aviation accident database Factual Reports. These Factual Reports were found to contain evidence of pilot behaviors that reflected hazardous attitudes and pilot actions that indicated poor ADM and CRM skills. This investigation utilized statistical hypotheses testing to illustrate the relationship between accident pilot behavior and accident pilot actions.

It is important to note that the pilot behavior and pilot action data for this investigation came from these NTSB Factual Reports and not the Probable Cause Reports. The Probable Cause Reports are an excellent summary of the primary causes and contributing factors involved in an accident. However, they do not address the wealth of additional information that can be found in the Factual Reports. Almost every Factual Report contained relevant information that was not included in the Probable Cause Report. For example, one of the accident pilots in our study, whose family indicated to an investigator that he might have been suffering from depression, left behind a suicide note before departing on his final flight. That important detail was not listed in the Probable Cause Report.

All pilots, whether they realize it or not, properly use, misuse, or do not use basic ADM and CRM skills on every flight. For example, before taking off, a pilot has to evaluate the risk involved with that particular flight. Then, a pilot has to make a series of aeronautical decisions that can affect the safe outcome of the flight. Throughout the flight, the pilot has to be alert to any potential piloting errors. Finally, a pilot has to know how to utilize, effectively, the proper cockpit resources to deal with any difficult situations that might arise during the flight. Pilots that make good use of ADM and CRM skills have an excellent chance of enjoying a successful flight. Pilots that either misuse or do not use these ADM and CRM skills are more likely to experience misfortune. This study will show that pilot behaviors (as evidenced by hazardous attitudes) have the potential to influence negatively pilot actions (as displayed by ADM and CRM skill utilization).

Literature Review

Federal Aviation Administration (FAA) Advisory Circular (AC) 60-22 provides a thorough discussion of ADM and hazardous attitudes (FAA, 1991). Additional information can be found on the application of ADM and hazardous attitudes regarding the flight training environment in the Aviation Instructor's Handbook (FAA, 1999) and the Instrument Flying Handbook (FAA, 2001).

Various researchers have published articles related to ADM. Case studies in hazardous attitudes are described by Kern (1998). The measurement of hazardous attitudes among pilots is discussed by Hunter (2005). Murray (1999) has proposed adding the fear of loss of face to the original list of hazardous attitudes.

FAA AC 120-51E provides detailed information on CRM (FAA, 2004a). Theoretical applications can be reviewed in Jensen (1995) and Fallucco (2002). Practical applications of CRM can be found in industry publications such as United Airlines (1996), Flight Safety International (1986), and Saudi Arabia Airlines (1993).

CRM continues to be a subject of scholarly study. Salas, Prince, and Bowers (1999) argue for more standardization in CRM training. Hedge et al (2000) discuss the methodology for selecting pilots with appropriate CRM skills. Pilot decision-making in instrument meteorological conditions are examined by Wiegmann, Goh and O'Hare (2002).

The Software, Hardware, Environmental, and Liveware (SHEL) model of system safety was first described by Edwards (1972). Other authors have provided additional insight on how this model can be applied to evaluating aviation risk factors (Wiener & Nagel, 1988; Hawkins, 1993; and, Wiener, Kanki & Helmreich, 1993).

Information on individual accidents can be found in the National Transportation Safety Board (NTSB) aviation accident databases (NTSB, 2005). Summaries of the aviation accident data are also available (NTSB, 2004; and, AOPA, 2002).

This study is the first paper from a comprehensive research project concerning ADM, CRM and flight training issues. Pilot age was found to have no effect on ADM and CRM skills including hazardous attitudes (Wetmore & Lu, 2005a). On the other hand, obtaining higher levels of pilot certification and more flight experience reduce the likelihood that a pilot will display hazardous attitudes (Wetmore & Lu, 2005b). Certain pedagogical paradigms were found to have either ameliorating or exacerbating effects on student pilots with hazardous attitudes (Wetmore & Lu, 2005c). Finally, flight instructor conflict management styles were found to have either a beneficial or a harmful effect on student pilots with hazardous attitudes (Wetmore & Lu, 2006).

Research Questions

The purpose of this paper is to study the effects of pilot behavior on pilot actions. More specifically, this study examines the relationship between hazardous attitudes and certain ADM and CRM skills. Some of the most important ADM and CRM skills are: (a) evaluation of risk; (b) decision-making; and (c) effective use of all available cockpit resources. The research questions, which fulfill this purpose, are defined as follows:

1. How do hazardous attitudes affect a pilot's risk-taking behavior?
2. How do hazardous attitudes affect a pilot's decision-making skills?
3. How do hazardous attitudes affect a pilot's ability to utilize all available cockpit resources?

Research Methods

A random number generator from Norusis (2004) was used to select at random 50 general aviation accidents that involved a fatality from the NTSB aviation accident databases (NTSB, 2005). Using a random number generator to select a sample from a targeted population should ensure the needed reliability for data analysis (Creswell, 2003).

The years 1997 to 2002 were selected because the vast majority of the Factual Reports had been upgraded from Preliminary to Final status. Factual Reports for each of the 50 accidents were downloaded from the NTSB website. These reports were then analyzed for evidence of ADM and CRM factors.

The first step of the analysis was to use the Factual Reports to assemble a chain of events for each accident. Both researchers had to agree on what constituted an event, the sequence of events, and the pilot actions associated with those events in regards to ADM and CRM. Next, the Factual Reports were examined for evidence of pilot behavior related to hazardous attitudes. Again, both researchers had to agree on what constituted a pilot behavior and if that behavior was reflective of a hazardous attitude.

During this endeavor, the researchers developed a complex set of guidelines to aid in hazardous attitude determination and classification. A detailed description of those guidelines is of such a length as to be beyond the scope of this paper. However, the authors are in the process of preparing a methodology and quantitative results paper that will present those guidelines and list the actual pilot behaviors uncovered in this study that were indicative of hazardous attitudes.

The researchers decided that this paper should focus on those relationships between pilot behaviors and pilot actions that can be examined using statistical measures. One-way Analysis of Variance (ANOVA) was used to test the null hypothesis that certain ADM and CRM population means were equal (Norusis, 2004). An observed significance level of 0.05 was used to reject the null hypothesis. An example of the null (Ho) and alternative (Ha) hypothesis testing is shown in Figure 1.

Ho & Ha for *(CRM Skill)* per Accident Flight and Hazardous Attitudes per Pilot

Ho: The *(CRM Skill)* mean for a pilot with hazardous attitudes is **equal** to the *(CRM Skill)* mean for a pilot without hazardous attitudes.

Ha: The *(CRM Skill)* mean for a pilot with hazardous attitudes is **not equal** to the *(CRM Skill)* mean for a pilot without hazardous attitudes.

Figure 1. Null Hypothesis Testing Example

The authors used logic and common sense to determine cause and effect in this study. For example, a pilot who cancels an instrument flight rules (IFR) clearance to fly below the clouds in mountainous terrain just made a bad aeronautical decision. It seems logical that the pilot's hazardous attitude of invulnerability contributed to that bad decision. Saying that the decision to cancel the IFR clearance caused the invulnerability would not make any sense.

To avoid circular logic, such as using the pilot action to determine the pilot behavior, there had to be collaborating evidence to support a hazardous attitude determination. In the example described above, the pilot elected to conduct the flight with an inoperative piece of navigation equipment on-board and failed to obtain an in-flight weather update before deciding to scud-run in the mountains. These three facts taken together are indicative of a pilot with feelings of invulnerability.

Results & Discussion

Accident Pilot Descriptive Statistics

The accident pilots in this study were 54 years old on average (see Table 1). Their total flight time had a mean of 3502 and a median of 1295 hours. The highest pilot certificate held by these pilots had a mean of 2.82 and a median of 2.00 certificates. The values for each certificate are as follows: (a) student pilot certificate =1; (b) private pilot certificate = 2; (c) commercial pilot certificate = 3; (d) certified flight instructor (CFI) certificate = 4; and (e) airline transport pilot (ATP) certificate = 5.

Table 1
Accident Pilot Descriptive Statistics

Descriptive Statistic	Mean	Median
Age (years)	54	54
Total Flight time (hours)	3502	1295
Highest Certificate Held (*)	2.82	2.00

* None = 0; Student = 1; Private = 2; Commercial = 3; CFI = 4; and, ATP = 5.

Accident Flight Descriptive Statistics

The accident flight descriptive statistics are shown in Table 2. A large majority of the accidents (84%) occurred during visual flight rule (VFR) conditions. Most of the accidents (88%) happened during the day. Over half of the accidents (60%) took place on cross-country flights.

Critical Flight Events

All of the cases in this study had a chain of events that eventually led up to the actual accident. Based on the Factual Reports (NTSB, 2005), all of these accident chains contained a critical event without which the accident probably would not have occurred. Not surprisingly, pilot error (58%) and mechanical failure (26%) were the most frequent critical events.

Table 2
Accident Flight Descriptive Statistics

Flight Rules	Condition	Location	Frequency	Percent (%)
VFR	Day	Local	19	38
VFR	Day	Cross-country	18	36
VFR	Night	Local	1	2
VFR	Night	Cross-country	4	8
IFR	Day	Cross-country	7	14
IFR	Night	Cross-country	1	2
Total Number of Flights			50	100

Table 3
Critical Flight Events

	Critical Flight Event	Frequency	Percent (%)
	Mechanical Failure	13	26
PILOT ERROR	IFR Spatial Disorientation	6	12
PILOT ERROR	VFR into IMC	5	10
PILOT ERROR	Surface Winds	4	8
PILOT ERROR	Mid-air Collision	3	6
PILOT ERROR	VFR CFIT	3	6
PILOT ERROR	Intentional Low Flight (buzzing)	2	4
PILOT ERROR	Fuel Exhaustion	2	4
PILOT ERROR	Thunderstorm	2	4
PILOT ERROR	Wake Turbulence	1	2
PILOT ERROR	Alcohol	1	2
	Hijacker	1	2
	Heart Attack	1	2
	Suicide	1	2
	Unknown	5	10
	Total Number of Flights	50	100

Frequency of ADM & CRM Factors

The Factual Reports (NTSB, 2005) were examined for evidence of causal or contributing ADM and CRM factors. The frequency of these causal or contributing ADM and CRM factors in the accident chains investigated are shown in Table 4. Certain ADM and CRM factors had a high frequency of occurrence: (a) taking risks; (b) displaying hazardous attitudes; (c) making poor aeronautical decisions; (d) committing pilot errors; and (e) failing to utilize all available cockpit resources.

Table 4

Frequency of ADM & CRM Factors in General Aviation Fatal Accidents

ADM & CRM Factors	Frequency	Percent (%)
Risk Factors	50	100
Hazardous Attitudes	43	86
Decision-making	42	84
Pilot Error	42	84
Cockpit Resource Utilization	40	80
Proficiency	34	68
Stress	15	30
Communication Skills	8	16
Disqualifying “Substance”	7	14
Cockpit Management	5	10
Currency	5	10
Physical Health	5	10
Teamwork	4	8
Personality Disorder	3	6
Leadership	2	4
Total Cases in this Study	50	

Frequency of Hazardous Attitudes

The types and frequencies of the hazardous attitudes evidenced by the accident pilots in this study are shown in Table 5. Invulnerability was the most common hazardous attitude. Resignation was the least common. The percentages total more than 100% because many of the accident pilots displayed more than one hazardous attitude. The multiple hazardous attitude frequency and cumulative percentages are shown in Table 6. The majority of the accident pilots (74%) evidenced two or more hazardous attitudes. Figure 2 is a histogram with a normal curve showing the normal distribution of multiple hazardous attitudes. The accident pilots evidenced a mean of 2.32 and a median of 3.00 hazardous attitudes per accident flight.

Table 5
Hazardous Attitude Types and Frequencies

Hazardous Attitude	Frequency	Percent (%)
Antiauthority	23	46
Impulsivity	24	48
Invulnerability	40	80
Macho	20	40
Resignation	9	18

Table 6
Hazardous Attitude Frequency and Cumulative Percent

Hazardous Attitudes: Number per Accident Flight	Frequency	Percent (%)	Cumulative Percent (%)
5	0	0	0
4	8	16	16
3	20	40	56
2	9	18	74
1	6	12	86
0	7	14	100
Total	50	100	

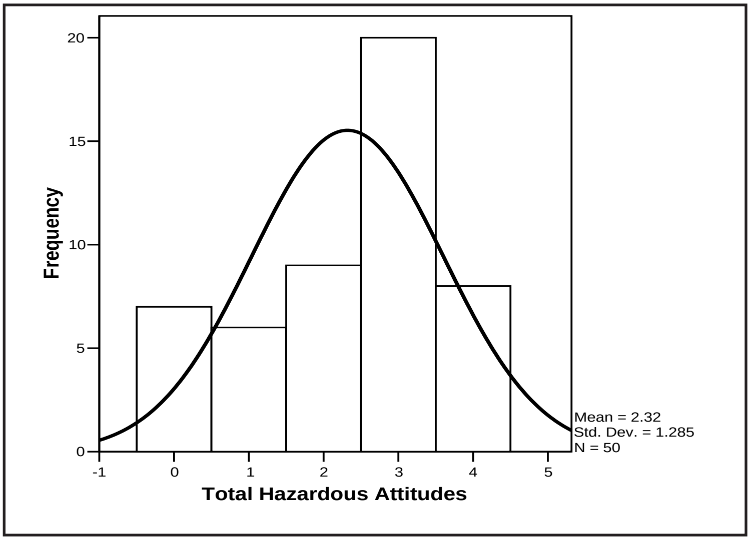


Figure 2. Histogram of Total Hazardous Attitudes per Accident Flight

Effects of Hazardous Attitudes on CRM Skills

Hazardous attitudes appear to affect certain accident pilot ADM and CRM skills (see Table 7). The effect of hazardous attitudes on specific ADM and CRM skills can be described as follows: (a) more willing to accept high-risk flights; (b) more prone to making bad decisions; (c) more likely to commit pilot errors; and (d) less likely to use all of the available cockpit resources. On the other hand, accident pilots with health issues appear to be less likely to exhibit hazardous attitudes. There was not enough evidence to reject the null hypothesis (see Figure 1) for the effects of hazardous attitudes on the other ADM and CRM skills listed in Table 7. Either hazardous attitudes do not affect these particular ADM and CRM skills or the affect is so small that it is undetectable.

Effects of Multiple Hazardous Attitudes on CRM Skills

Evidence for multiple hazardous attitudes in the accident pilots of this study had a measurable effect on some of their ADM and CRM skills (see Table 8). Comparing pilots with zero hazardous attitudes to pilots with four hazardous attitudes indicates the following: (a) nearly doubles the number of risk factors accepted per accident flight; (b) almost quintuples the number of bad decisions made per accident flight; and (c) goes from full utilization of cockpit resources to not using two available resources per accident flight. Accident pilots who are flying with health issues appear less likely to exhibit multiple hazardous attitudes. There was not enough evidence to reject the null hypothesis (see Figure 1) for the effects of multiple hazardous attitudes on the other ADM and CRM skills listed in Table 8. Either hazardous attitudes do not affect these particular CRM skills or the affect is so small that it is undetectable.

Table 7

Effects of Hazardous Attitudes on CRM Skills

	Hazardous Attitudes (Haz Att)			
	Average Occurrences per Flight			
CRM Skills	With Haz Att	Without Haz Att	Significance (0.05)	Haz Att Effect on CRM Skill
Risk Factors	8.42	5.71	0.026	Increase
Bad Decisions	1.95	0.57	0.011	Increase
Pilot Errors	1.21	0.57	0.028	Increase
Resources not Utilized	2.16	0.00	0.000	Increase
Lack of Pilot Proficiency	1.16	1.29	0.791	*
Pilot Stressors	0.30	0.29	0.931	*
Poor Communications	0.19	0.00	0.221	*
Disqualifying "Substances"	0.12	0.29	0.239	*
Cockpit Management Issues	0.12	0.00	0.352	*
Lack of Pilot Currency	0.12	0.00	0.352	*
Pilot Physical Health Issues	0.05	0.43	0.001	Decrease
Teamwork Issues	0.07	0.14	0.518	*
Personality Disorders	0.07	0.00	0.481	*
Leadership Issues	0.05	0.00	0.570	*

* Not enough evidence to reject the null hypothesis. Either hazardous attitudes do not affect this CRM skill or the affect is so small as to be undetectable.

Table 8
Effects of Multiple Hazardous Attitudes on CRM Skills

	Multiple Hazardous Attitudes (Mult Haz Att)						
	Number of Haz Att per Pilot and Average Occurrences per Flight						
CRM Skill	None	One	Two	Three	Four	Significance (0.05)	Mult Haz Att Effect on CRM Skill
Risk Factors	5.71	8.17	6.89	8.45	10.25	0.031	Increase
Bad Decisions	0.57	1.00	1.44	2.10	2.88	0.003	Increase
Pilot Errors	0.57	0.83	1.33	1.25	1.25	0.146	*
Resources not Utilized	0.00	2.17	1.89	2.35	2.00	0.001	Increase
Lack of Pilot Proficiency	1.29	1.17	1.56	1.00	1.13	0.820	*
Pilot Stressors	0.29	0.50	0.11	0.30	0.38	0.602	*
Poor Communications	0.00	0.00	0.33	0.10	0.38	0.100	*
Disqualifying “Substances”	0.29	0.00	0.11	0.15	0.13	0.704	*
Cockpit Management Issues	0.00	0.00	0.11	0.05	0.38	0.066	*
Lack of Pilot Currency	0.00	0.17	0.11	0.10	0.13	0.900	*
Pilot Physical Health Issues	0.43	0.00	0.00	0.10	0.00	0.023	Decrease
Teamwork Issues	0.14	0.17	0.00	0.05	0.13	0.717	*
Personality Disorders	0.00	0.17	0.00	0.10	0.00	0.533	*
Leadership Issues	0.00	0.00	0.11	0.00	0.13	0.420	*

* Not enough evidence to reject the null hypothesis. Either hazardous attitudes do not affect this CRM skill or the affect is so small as to be undetectable.

How do Hazardous Attitudes Affect Risk Taking?

Software, hardware, environmental and/or liveware risk factors were involved in each of the accident flights (see Table 4). A histogram with a normal curve showing the normal distribution of multiple risk factors per accident flight is shown in Figure 3. All of the accident flights evidenced multiple risk factors. The number of risk factors per accident flight ranged from a minimum of three to a maximum of 13. There was a mean of 8.04 and a median of 8.00 risk factors per accident flight.

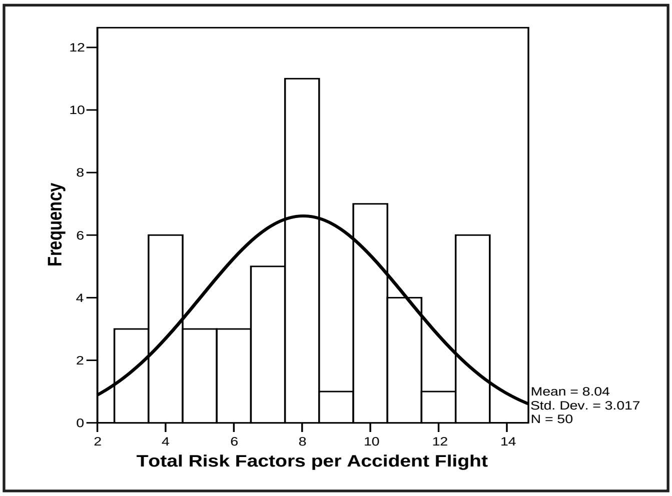


Figure 3. Histogram of Total Risk Factors per Accident Flight

It is likely that accident pilots with hazardous attitudes are more willing to accept flights with increased risk factors (see Table 7). Accident pilots with hazardous attitudes accepted an average of 8.42 risk factors per accident flight while those accident pilots without hazardous attitudes accepted an average of 5.71 risk factors per accident flight.

It is likely that the more hazardous attitudes that an accident pilot exhibits, the more likely it is for that pilot to accept flights with multiple risk factors (see Table 8). An accident pilot with four hazardous attitudes accepted an average of 10.25 risk factors per accident flight while an accident pilot without any hazardous attitudes accepted an average of 5.71 risk factors per accident flight (see Figure 4).

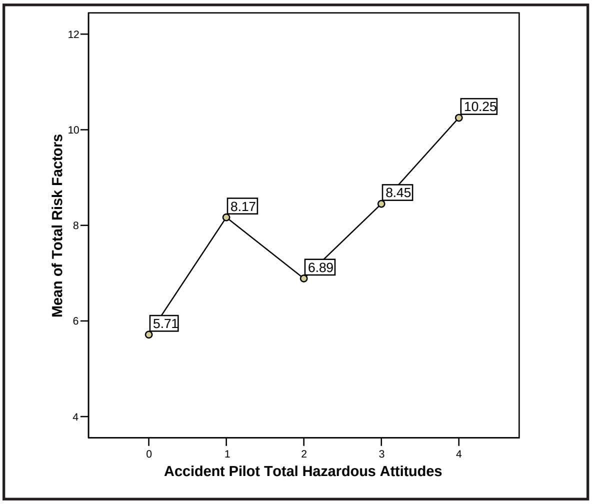


Figure 4. Relationship between Multiple Hazardous Attitudes and Risk Taking Behavior per Accident Flight

How do Hazardous Attitudes Affect Decision-making?

A histogram with a normal curve showing the normal distribution of multiple bad decisions per accident flight is shown in Figure 5. The accident pilots had a mean of 1.76 and a median of 1.00 bad decisions per accident flight. Almost half (48%) of the accident pilots made two or more poor aeronautical decisions per accident flight.

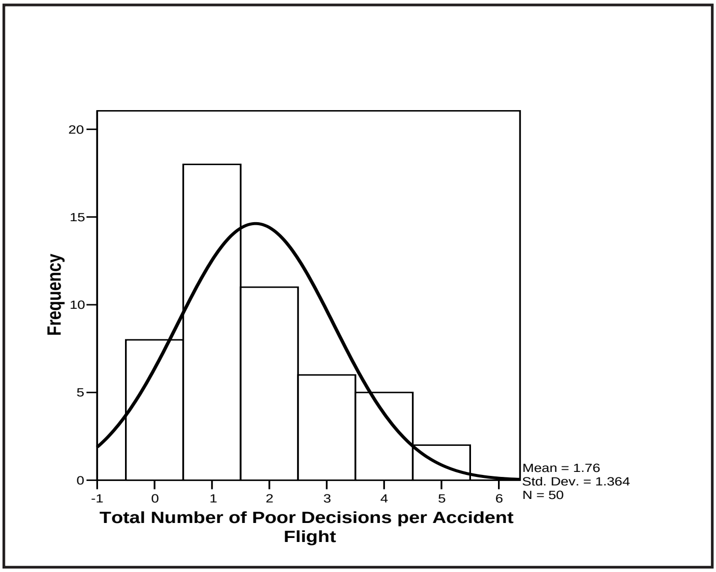


Figure 5. Histogram of Poor Aeronautical Decision-making per Accident Flight

It is likely that accident pilots with hazardous attitudes are more prone to make bad decisions (see Table 7). On average, accident pilots with hazardous attitudes made 1.95 poor decisions per accident flight. Those accident pilots without hazardous attitudes made an average of 0.57 poor decisions per accident flight.

Pilots with multiple hazardous attitudes are more likely to make multiple bad decisions during an accident flight (see Table 8). The relationship between multiple hazardous attitudes and aeronautical decision-making (see Figure 6) can be summarized as follows: (a) pilots with one hazardous attitude averaged almost twice as many bad aeronautical decisions as pilots without hazardous attitudes; (b) pilots with two hazardous attitudes averaged about two and a half (2 ½) times as many bad aeronautical decisions as pilots without hazardous attitudes; (c) pilots with three hazardous attitudes averaged more than three and a half (3 ½) times as many bad aeronautical decisions as pilots without hazardous attitudes; and (d) pilots who evidenced four hazardous attitudes averaged approximately five (5) times as many bad aeronautical decisions as pilots without hazardous attitudes

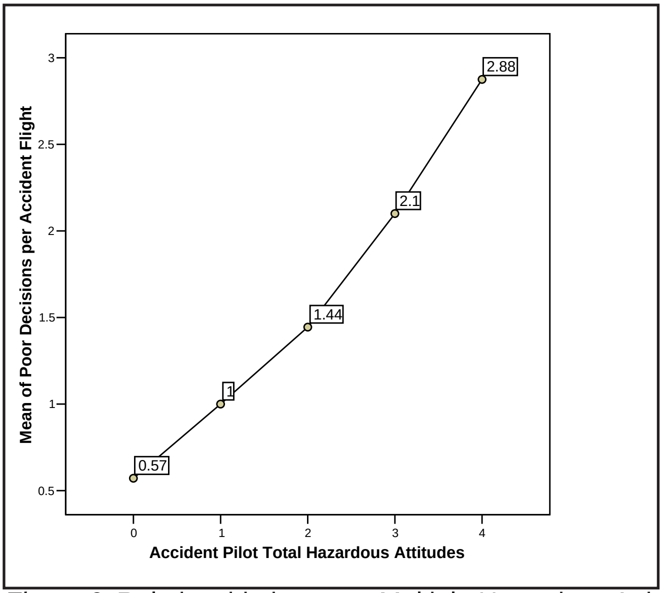


Figure 6. Relationship between Multiple Hazardous Attitudes and Aeronautical Decision-making per Accident Flight

How do Hazardous Attitudes Affect the Utilization of Cockpit Resources?

Accident pilots with hazardous attitudes appear to be less likely to use all available cockpit resources (see Table 7). Accident pilots with hazardous attitudes failed to use an average of 2.16 cockpit resources per accident flight. Accident pilots without hazardous attitudes did not display any evidence that they failed to use all available cockpit resources per accident flight.

A histogram with a normal curve showing the normal distribution of cockpit resources not used per accident flight is shown in Figure 7. The accident pilots had a mean of 1.86 and a median of 2.00 cockpit resources not used per accident flight. More than half (56%) of the accident pilots failed to utilize two or more available cockpit resources.

Accident pilots with multiple hazardous attitudes are more likely not to utilize multiple cockpit resources during an accident flight (see Table 8). The effect of multiple hazardous attitudes on cockpit resource utilization is shown in Figure 8.

Conclusions

Hazardous attitudes lead to risk-taking behavior. Accident pilots with hazardous attitudes averaged 8.42 risk factors per accident flight (see Table 7). Accident pilots without hazardous attitudes averaged only 5.71 risk factors per accident flight. Multiple hazardous attitudes lead to increased risk-taking behavior (see Table 8). Accident pilots who evidenced four hazardous attitudes averaged 10.25 risk factors per accident flight.

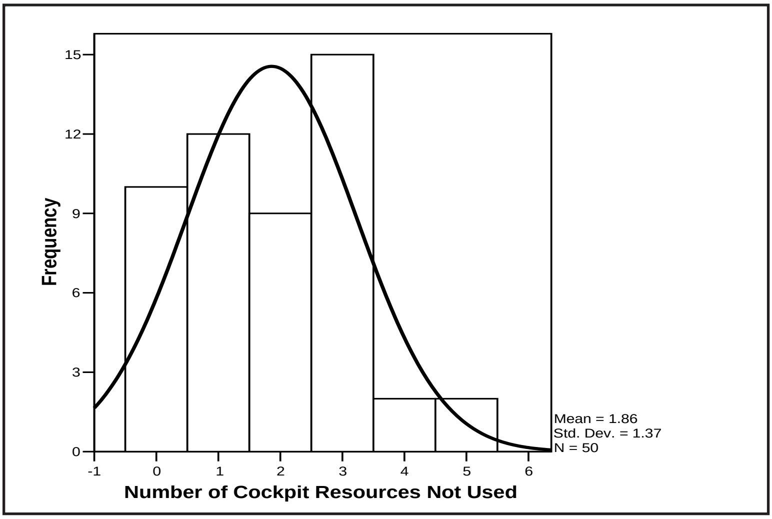


Figure 7. Histogram of Cockpit Resources Not Utilized per Accident Flight

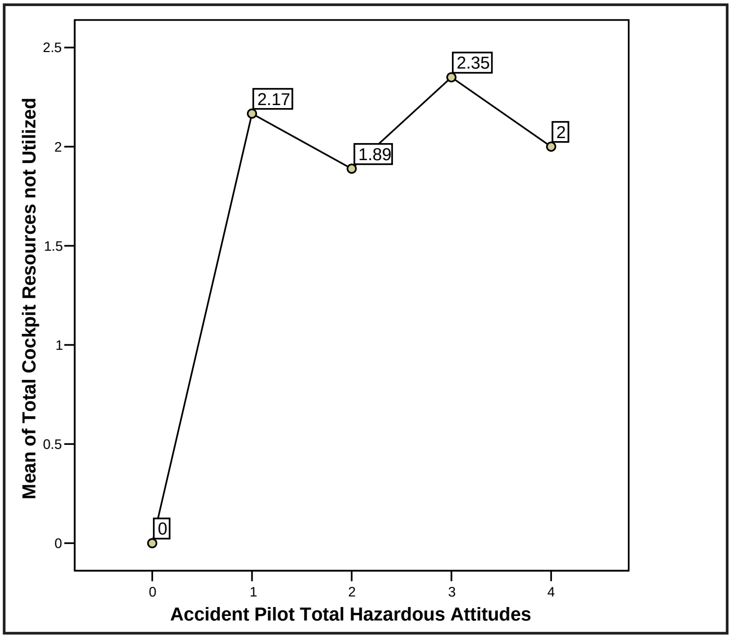


Figure 8. Relationship between Multiple Hazardous Attitudes and Cockpit Resource Utilization per Accident Flight

Hazardous attitudes affect decision-making skills. Accident pilots with hazardous attitudes made an average of 1.95 bad decisions per accident flight (see Table 7). Accident pilots without hazardous attitudes made an average of 0.57 bad decisions per accident flight. Multiple hazardous attitudes result in an increased number of bad decisions (see Table 8). Accident pilots with four hazardous attitudes averaged 2.88 bad decisions per accident flight.

Hazardous attitudes affect a pilot's ability to utilize effectively all available cockpit resources (see Table 7). Accident pilots with hazardous attitudes failed to utilize 2.16 cockpit resources per accident flight. Accident pilots without hazardous attitudes did not display any evidence that they failed to utilize all available cockpit resources per accident flight. Multiple hazardous attitudes exacerbate this tendency of not using all available resources (see Table 8).

Those accident pilots with hazardous attitudes were more likely to commit pilot errors (see Table 7). However, the effect of multiple hazardous attitudes on multiple pilot errors is not significant (see Table 8). Either multiple hazardous do not affect the number of errors committed by a pilot, or the affect is so small that it cannot be detected. The explanation for this observation is open to debate. One possible explanation is that in many cases the accident aircraft crashed after one pilot error. In other words, the accident pilots just did not have the opportunity to commit multiple errors.

Hazardous attitudes have a devastating effect on three of the most important CRM skills: (a) evaluating risk, (b) making decisions, and (c) utilizing all available resources. Based on the Factual Reports (NTSB, 2005), this is how a typical accident scenario involving hazardous attitudes unfolds: (a) hazardous attitudes influence a pilot into accepting a high-risk flight; (b) when the flight situation begins to deteriorate, hazardous attitudes contribute to the pilot making bad decisions; (c) when this high-risk flight encounters a critical event, and the pilot is making bad decisions, hazardous attitudes prevent that pilot from utilizing all of the available cockpit resources and thus changing the tragic outcome of the flight.

Our study suggests that general aviation pilots involved in fatal accidents were deficient at certain ADM and CRM skills. What is the solution to this problem? One possible answer is that all pilots, at every level of pilot certification, should receive better ADM and CRM training. The practical test standards (PTS) already list ADM and CRM as special emphasis areas for all pilot certifications and ratings (FAA, 2002a; FAA, 2002b; FAA, 2002c; and FAA, 2004b). We recommend that ground instructors, flight instructors, stage check instructors, aviation educators, and pilot examiners carefully read these PTS special emphasis areas in order to create innovative flight training and/or case-based scenarios that stress the importance of ADM and CRM skills in the flight training environment.

Future Study

The data generated by this study is quite fascinating. The authors plan to investigate the effects of age, experience, and training on CRM skills. The authors also intend to examine the effects of each of the individual hazardous attitudes on selected CRM skills such as flight proficiency, cockpit management, and stressors. A paper on how pilot's can recognize hazardous attitudes in themselves and others could also be beneficial to the aviation community. Finally, an interest has been expressed in setting forth the complex and detailed guidelines that we used in this current study to gather evidence for ADM and CRM factors.

This current inquiry is focused on Part 91 operations. Our logic for this as a starting point is that almost all pilots begin their flying careers in general aviation. Our long-range plans are to conduct similar investigations on Part 135 and Part 121 operations. The culmination of these long-range plans is to look for correlations among the different levels of professionalism. The goal of this research is to promote a safer flying environment for all pilots.

References

Aircraft Owners and Pilots Association Air Safety Foundation. (2002). *2002 Nall report: accident trends and factors for 2001*. Author.

Creswell, J. W. (2003). *Research design: qualitative quantitative and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Edwards, E. (1972). Man and machine: Systems for safety. *Proceedings of British Airline Pilots Association Technical Symposium* (21-36). London: British Airline Pilots Association.

Fallucco, S. J. (2002). *Aircraft command techniques*. Burlington, VT: Ashgate Publishing Company.

Federal Aviation Administration. (1991). *Aeronautical decision making*. (FAA Advisory Circular 60-22). Author.

Federal Aviation Administration. (1999). *Aviation instructor's handbook* (FAA-H-8083-9). Author.

Federal Aviation Administration. (2001). *Instrument flying handbook* (FAA-H-8083-15). Author.

Federal Aviation Administration. (2002a). *Private pilot practical test standards* (FAA-S-8081-14AS). Author.

Federal Aviation Administration. (2002b). *Commercial pilot practical test standards* (FAA-S-8081-12B). Author.

Federal Aviation Administration. (2002c). *Flight instructor practical test standards* (FAA-S-8081-6BS). Author.

Federal Aviation Administration. (2004a). *Crew resource management training* (FAA Advisory Circular 120-51E). Author.

Federal Aviation Administration. (2004b). *Instrument rating practical test standards* (FAA-S-8081-4D). Author.

Flight Safety International. (1986). *Practical cockpit management*. Author.

Hawkins, F. H. (1993). *Human factors in flight* (2nd ed.). Brookfield, VT: Ashgate Publishing Company.

Hedge, J. W., Bruskiewicz, K. T., Borman, W. C., Hanson, M. A., & Logan, K. K. (2000). Selecting pilots with crew resource management skills. *The International Journal of Aviation Psychology*, *10*(4), 377-392.

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

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

Kern, T. (1998). *Flight discipline*. New York: McGraw-Hill.

Murray, S. R. (1999). FACE: Fear of loss of face and the five hazardous attitudes concept. *The International Journal of Aviation Psychology*, *9*(4), 403-411.

National Transportation Safety Board. (2004). *Annual review of aircraft accident data U.S. general aviation, calendar year 2000* (NTSB/ARG-04/01). Author.

National Transportation Safety Board. (2005). Data and information products. Aviation accident database. Retrieved July 19, 2005, from http://www.ntsb.gov/ntsb/query.asp.

Norusis, M. J. (2004). *SPSS 12.0 guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.

Salas, E., Prince, C., & Bowers, C. A. (1999). A methodology for enhancing crew resource management training. *Human Factors*, *41*(1), 161-172.

Saudi Arabian Airlines. (1993). *Crew resource management*. Author.

United Airlines. (1996). *Command/leadership/resource management*. Author.

Wetmore, M. J., & Lu, C-t. (2005a). The effects of pilot age on aeronautical decision making and crew resource management skills. *White paper*. Central Missouri State University.

Wetmore, M. J., & Lu, C-t. (2005b). Reducing hazardous attitudes: the effects of pilot certification and flight experience. *White paper*. Central Missouri State University.

Wetmore, M. J., & Lu, C-t. (2005c). The effects of pedagogical paradigms on aviation students with hazardous attitudes. *White paper*. Central Missouri State University.

Wetmore, M. J., & Lu, C-t. (2006). Conflict management strategies related to hazardous attitudes in the flight training environment. White paper. Central Missouri State University.

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilot's decisions to continue visual flight rules flight into adverse weather. Human Factors, 44(2), 189-197.

Wiener, E. L., & Nagel, D. C. (1988). Human factors in aviation. London: Academic Press.

Wiener, E. L., Kanki, B. G. & Helmreich, R. L. (1993). Cockpit resource management. London: Academic Press.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Identifying Effective Leader Behavior and Its Relationship to the Exercise of Power in the Cockpit: A Multidimensional Scaling Study

Raymond Allen Hamilton II
Assistant Professor of Aviation Management
College of Business
Auburn University, Alabama 36849-5266

Abstract
Presented is a classification of effective pilot-in-command behaviors using multidimensional scaling methodology. In this context, effective behavior is defined as that which is conducive to team building and solidarity in the air carrier cockpit. To cover the full spectrum of the construct, a list of 24 behavioral descriptors were derived from prior research. A survey soliciting constructive, team-building leader behaviors in the cockpit was then administered to 48 flight officers and 29 flight attendants employed by a major air carrier. Based on this pool of items, a decompositional method, multiple dimensional scaling, was applied to determine a classification of constructive leader behaviors and their relationship to the exercise of power. The derived classification identified the underlying dimensions, which will hopefully contribute to theory development and direction for further research in improving pilot selection methods.

Leadership behavior differs from leadership skill in that behavior relates to action in contrast to leadership skill that relates to knowing when to act and acting in a way suitable to the circumstance and in a manner conducive to goal accomplishment. This focuses on those behaviors that color the perceptions of what constitutes an effective team leader in the cockpit and how those behaviors relate to the exercise of power. Every commercial flight crew constitutes a team (Helmrich, 1984; Helmrich & Wilhelm, 1989). How well it functions as a team depends

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.

on three factors related to crew resource management (CRM) training (Orlady & Orlady, 1999):

1. Crewmembers' skill as team members. One cannot assume that these skills exist. Flight training does not address team skills as a centerpiece. The average U.S. flight school trains pilots to function solo--not as a member of a well-coordinated flight crew.
2. The criteria airlines employ in screening and selecting pilot hires. The focus of the selection process is almost exclusively on technical and decision-making skill along with psychopathological and physiological screening—this to the exclusion of any assessment of an applicant's capacity and proclivity to function well as a team member.
3. The airline system of crew assignment and scheduling. Crew assignments are changed monthly and are based primarily on a seniority-controlled queuing system. The result precludes the exposure and experience needed for a crew to bond and develop as a well coordinated team, hence the criticality of the flight captain's initial actions as a team leader (Hackman, 1993).

The purpose of the research is to explore flight crew perceptions of effective team leadership exhibited by line-assigned flight captains. Second officers on probationary status were chosen as subjects since they are most dependent on the flight captain's leadership style and approach to the exercise of social power over the crew. Additionally, the dynamics of a three- versus a two-person flight crew provided greater demand for coordination and effective team leadership, hence greater opportunity to observe leadership style and approach to the exercise of power. With knowledge of these perceptions, the goal is to gain an understanding of the flight captain's image as an effective team leader and what behaviors contribute to the development of rapport and trust—results that the pilot-in-command cannot decree.

Method

Perceptual mapping was employed to examine data gathered in a series of interviews with 2nd officers in their first year of probation with a major air carrier. Perceptual mapping or multidimensional scaling (MDS) is a procedure that allows the researcher to identify the perceived relative image of a set of objects related to shared perceptions among a group of subjects or participants (Green & Rao, 1972; Kruskal & Wish, 1978; Schiffman, Reynolds, & Young, 1981). In this instance, the objects consist of behavioral descriptors related to the actions of effective team leadership as observed by the participants (2nd officers) in the study. The objective was to transform crewmember (2nd officer) judgments of similarity regarding behavioral descriptors into distances represented in multidimensional space. Given two descriptors judged by the participants to be the most similar compared with all other combinations of descriptors, perceptual mapping will position the two descriptors so that the distance between them in multidimensional space is smaller that the distance between any other descriptor pairs. The derived perceptual or spatial map shows the relative positioning of all behavioral descriptors. MDS is based on comparison of objects; hence, any object can be viewed as possessing both perceived (subjective) and objective dimensions (Schiffman, Reynolds & Young, 1981). For example, a researcher may perceive one descriptor as crucial to the perception of a flight captain as a highly effective team leader whereas, a second officer on probation may not share that perception at all. Two descriptors may have the same characteristics of observed behavior but be viewed differently because the two behaviors are perceived to differ in intensity or relevance to effective team leadership. As a result, the dimensions perceived by the 2nd officers might not correlate with the objective dimensions the research envisaged. The subjects may possibly endorse different sets of objective characteristics and even adjust the importance they assign to each dimension. In addition, the subjects' evaluation of the dimensions may not be independent and may not agree even given the circumstance where the perceived dimensions coincide with the objective dimensions. The challenge was first to understand the perceived dimensions with the view to possibly relate them to the objective dimensions at a later stage of the research. This follow-on stage sought to assess how social power (French and Raven, 1959) might predict the position of each descriptor in both perceptual and objective space.

Derivation of the Descriptors

To cover the full spectrum of the construct, a list of 24 behavioral descriptors was derived from prior research. A survey was then administered to 48 flight officers and 29 flight attendants employed by the same carrier as the group of 2nd officers. To avoid bias, none of the subjects participated in the survey. This group was tasked to reduce the 24 behavioral descriptors to the ten most indicative of effective team leadership of an aircrew. Commonality of view ranged from 98.7 to 75.3 percent among the top ten descriptors shown below in random order:

1. Makes everyone enthusiastic about flying with him/her.
2. I trust his/her capacity to handle any in-flight emergency or security threat.
3. Content in letting me continue my first officer duties in the way I see fit.
4. Is satisfied with my performance as long as I don't present any problems.
5. Does not try to change anything as long as things are progressing well as a crew.
6. Shows me how to stay out of trouble with the pilot union and the airline.
7. Talks a lot about the "lay of the land" and what it takes to get what you want from the company (airline).
8. Lets me know when I am doing my job well.
9. Tells me what to do if I want to be rewarded for my efforts.
10. His/her ideas have stimulated me to reexamine my flying techniques and crew management, which I never before questioned.

Discussion

MDS is based on comparison of objects; hence, any object can be viewed as possessing both perceived (subjective) and objective dimensions (Green & Rao, 1972). Data was gathered by administering a survey instrument to 2nd officers employed by a major air carrier. In the perceptual mapping analysis a decompositional methodology associated with multidimensional scaling was employed whereby the subject (2nd officer) provides an overall evaluation of similarity or preference between objects (Levine, 1979). Using the 9.0 version of StataCorp statistical software, this set of evaluations was then "decomposed" or reduced into a set of “dimensions” or implicit factors that represent the overall differences among the behavioral descriptors (StataCorp, 2005).

Objectives of Perceptual Mapping

Examined was the importance of team member endorsement of the leader in the cockpit and how that related to observable behavior. The data were analyzed in two phases consisting of (1) identifying the position of behavioral descriptor 1 (Makes everyone enthusiastic about flying with him/her) in a perceptual map of the objects (behavioral descriptors) with an understanding of the dimensions comparison used by the subjects (2nd officers), and (2) assessment of the preferences toward descriptor #1 relative to the other objects. Before addressing the results, the process of data collection is shown below:

Perceptual Mapping Study

This study is based on surveys and interviews with 18 second officers on probation in there first year of hire with a major air carrier. The ten behavioral descriptors represent the basic behaviors of the flight captains that were deemed effective team leaders in the cockpit. Three types of data were collected: similarity judgments, rating of descriptors, and preferences for each descriptor in various operational conditions in flight.

Similarity Data. Data collection began with obtaining the perceptions of the subjects regarding the similarity or dissimilarity of the 10 behavioral descriptors. Similarity judgments were made with the comparison-of-paired-objects approach. The 45 pairs of descriptors were presented to the subjects, who determined the degree of similarity of each on a nine-point scale (1 being “not at all similar” and 9 being “very similar”). These results were transformed since increasing values for the ratings are based on similarity, the opposite of a distance measure.

Table 1
Behavioral Descriptors’ Similarity Ratings

	Descriptor									
Descriptor	1	2	3	4	5	6	7	8	9	10
1										
2	6.56									
3	6.01	5.41								
4	2.45	2.59	3.41							
5	2.76	2.48	4.02	7.01						
6	3.98	2.41	2.24	3.98	2.41					
7	2.46	3.46	3.92	2.29	2.24	3.97				
8	2.45	2.41	3.68	2.63	2.57	3.63	2.30			
9	2.53	5.01	6.68	2.46	7.11	5.68	2.79	2.52		
10	6.24	7.01	2.79	2.46	2.46	3.46	7.01	2.53	2.41	

Similarity/Dissimilarity Ratings for Each Descriptor										
Similarities > 6.0	2, 10	1, 10	9	5	4, 9	0	10	0	3, 5	1, 2, 7
Greatest Dissimilari-ties	4, 8	6, 8	6	7	7	3	4	7	10	9
Explanation: Similarity ratings are on a nine-point scale (1 = not at all similar, 9 = very similar)										

Attribute Ratings. In addition to the similarity judgments, Ratings of each descriptor were obtained for five categories of social power by two methods. The categories included referent, expert, legitimate, reward, and coercive power (French and Raven, 1959). In the first method, each object or descriptor was rated on a six-point scale for each category. In the second method, each subject was asked to select the descriptors best characterized by each category. A "pick-any-method" was used whereby the subject could pick any number of descriptors for each category (Levine, 1979).

Assumptions regarding Perceptual Mapping

The focus was on the comparability and representativeness of the descriptors under evaluations and the subjects. Regarding the sample of 18 second officers, a representative sample was sought by using a random sampling procedure coordinated through the new hire probation monitor at the major air carrier's flight operations. All of the descriptors evaluated in the perceptual mapping were endorsed by as least seventy-five percent of the 77 active crew members solicited for behavioral descriptions. This insured that positioning discrepancies would be attributed to the perceptual differences among the subjects (2nd officers).

Multidimensional Scaling

The next step was to determine the proper dimensionality and to show the results in a perceptual map. The indices of fit at each dimensionality were assessed. Table 2 contains the indices of fit for solutions of two to five dimensions.

Table 2
Assessing Model Fit and Dimensionality

Average Measures of Fit [a]				
Solution Dimensions	Stress [b]	Percentage Change	R Squared [c]	Percentage Change
5	.20066	-	.6307	-
4	.21354	6.4	.5559	11.9
3	.23658	10.8	.5009	9.9
2	.30039	30.0	.3934	21.5
[a] Average of 18 separate solutions				
[b] Kruskal's stress formula				
[c] Proportion of orginal similarity ratings determined by distances from the perceptual map				

As the table reflects, significant improvement was observed in expanding from two to three dimensions, after which the improvement drops off and holds constant as the number of dimensions increase. Weighing this gain in fit versus the challenge of interpreting three or more dimensions given 10 objects, two dimensions were chosen for further analyses. The perceptual map reflecting two dimensions is shown in Figure 1.

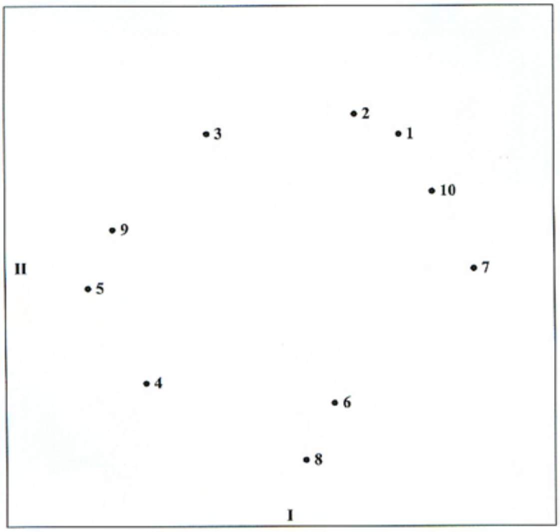


Figure 1. Perceptual Map of the 10 Behavioral Descriptors

Descriptor 1 is most closely associated with descriptor 2 with subjects viewing them virtually the same. Other pairs considered very similar are descriptors 6 and 8, 5 and 9, and 7 and 10 respectively. Descriptor 1 also differs from 4, 6, and 8 relative to dimension II, while dimension I distinguishes descriptor 1 from descriptors 3, 4,5, and 9 in one direction and descriptors 7 and 10 in another direction. These differences are viewed in their relative placement on the perceptual map. The map allows comparisons among all sets of descriptors. The challenge was interpreting the dimensions themselves. The solution fit was examined in a scatter plot of scaled similarity value versus fitted distances from the perceptual map (figure 2).

Examining a scatter plot of scaled similarity values versus distances on the perceptual map failed to reveal any significant outliers; thus, none of our descriptors and subjects was subject to elimination from our analysis. Kruskal's stress formula (Kruskal & Wish`, 1978) was calculated; the average among the 18 subjects was .296 and an R2 .402 based on the proportion of original similarity ratings accounted for by the distances extracted from the perceptual map. All subjects were shown on the perceptual map with the lowest measure of .31. Hence, no subject was a candidate of elimination. Table 1 reflects as well the high similarities (>.6) and the lowest similarity for each descriptor; hence, the primary patterns are easily recognized and compared to the perceptual map.

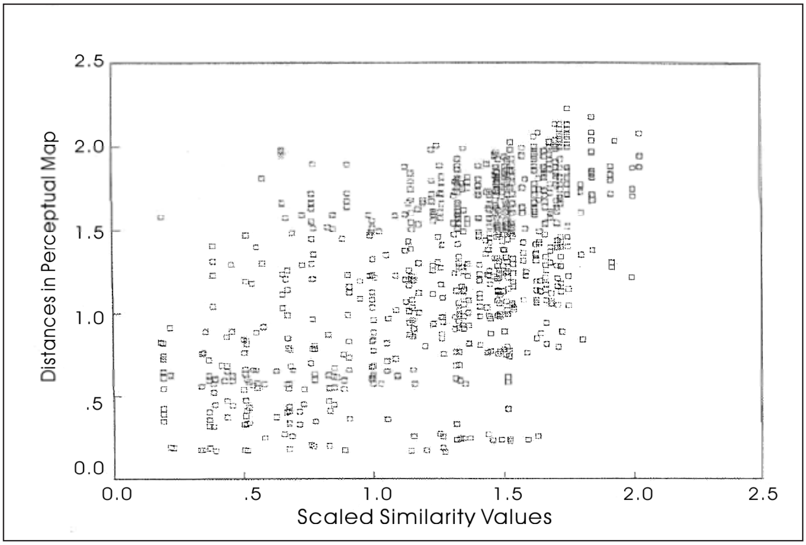


Figure 2. Scatter Plot

Interpreting the Results

The perceptual map presented the overall similarity on the subjects' judgments; however, to successfully interpret the results, were collected on five categories of social power: referent, expert, legitimate, coercive, and reward power (French and Raven, 1959). The ratings were averaged for an overall representative rating. To aid in interpreting the ratings, a vector model was used to match the ratings to the descriptor positions on the map (Figure 3). There emerged two dimensions of descriptors.

The first reflects reward power and expert power oriented opposite to expert power. This difference indicates a negative correspondence of expert versus reward power. The second group of descriptors indicates a close relationship of legitimate and coercive power pointed in the same direction but in opposition to referent power. To interpret the dimensions, the categories of power and their relative alignment with the axis were analyzed. Note that the vectors are somewhat angled from the axis. Since the map is a point representation, the vectors were rotated without affecting the relative positions. By rotating the axis about 40 degrees clockwise, an option routinely exercised in factor analysis, a dimension of expert versus reward power and a second dimension of referent versus legitimate/coercive power were identified. The rotation is not needed since the relative position of the descriptors to the resultant dimensions is clearly discernible. Table 3 shows the projection from each descriptor to the power vector as well as the original ratings; this aided in judging to what degree each vector represents the subject perceptions.

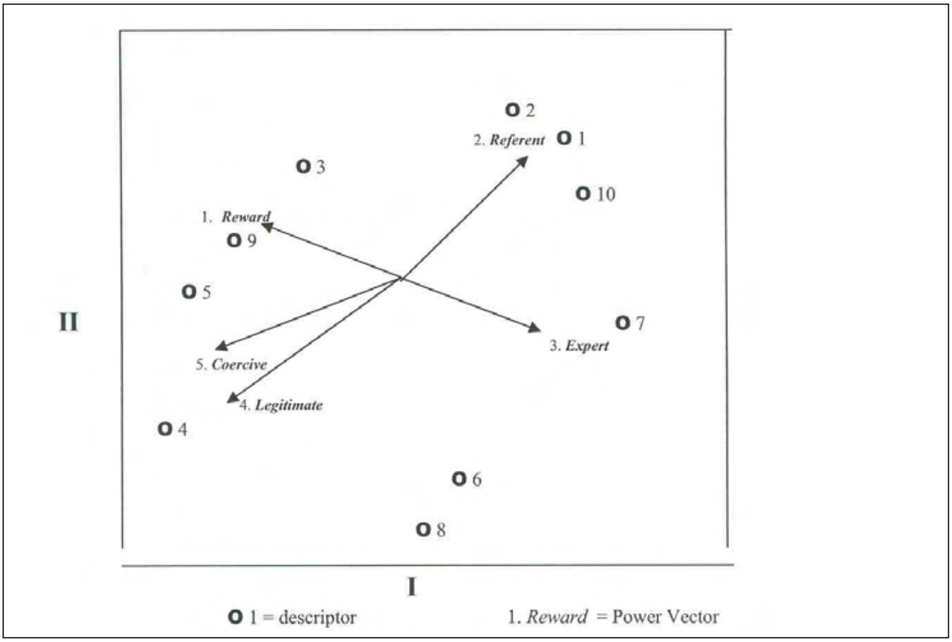


Figure 3. Integration of Behavioral Descriptors and Power Vectors on the Perceptual Map

Table 3
Descriptors Ratings vs. Projections on the Power Vectors

	Behavioral Descriptors										
Categories	1	2	3	4	5	6	7	8	9	10	Fit [a]
1. Reward	5.22 -.568	3.47 -.297	6.41 .672	5.88 1.132	6.06 1.490	4.94 -.556	5.29 -1.464	4.82 -.388	8.35 1.200	4.65 -1.233	.661
2. Referent	7.01 1.010	7.17 1.530	7.67 1.106	3.22 -.628	4.78 .247	5.11 -1.413	6.56 -.510	1.61 -1.682	8.78 .572	3.17 .151	.719
3. Expert	6.94 .199	5.78 -.070	3.41 -.910	3.67 -.808	3.67 -1.370	6.94 .888	6.44 1.422	7.22 .818	4.94 -1.209	6.11 1.032	.848

4. Legitimate	4.0 -1.291	1.83 -1.164	6.33 -.229	7.72 1.535	6.09 1.189	5.78 .574	5.50 -.964	6.11 .937	7.50 .672	4.17 -1.260	.789
5. Coercive	4.17 -1.106	1.56 -.896	6.06 .133	8.22 1.496	7.75 1.420	4.28 .144	3.89 -1.259	6.33 .455	7.72 .958	5.06 -1.356	
The upper values in each cell are the original descriptor ratings; lower numbers are the projections for the fitted vectors.											
[a] Fit measures the degree to which the descriptor ratings correlate with the vector projections											

There is a close correlation between the ratings and the vector projections. A measure of fit statistically for each descriptor shows how well the ratings correlate with the vector projections. For example, Descriptor 1, "Makes everyone enthusiastic about flying with him/her," has a correlation of .701 with the referent power vector. Since the perceptual map is predicated on an overall evaluation (which may not directly compare to the ratings) and the ratings themselves are averaged across subjects with values dependent on differences between subjects as well as descriptors, an exact fit would not be possible. Even so, the level of fit for the power categories separately and together is more than acceptable. These results support the need for the team leader in the cockpit, the flight captain, to establish rapport with the flight crew as well as the cabin crew. Note that the descriptor #2, "I trust his/her capacity to handle any in-flight emergency or security threat," shows a stronger relationship to referent power than expert power. Technical knowledge and flying skills only address secondarily what is in the minds of crew members when they consider the flight captain's anticipated performance in the leading the team during an emergency (Orlady, 1999). The command authority vested in the flight captain by the company arguably is inconsequential when considering how the flight captain is perceived as a leader when faced with a crisis. Although these results are neither comprehensive nor conclusive, they do strongly suggest that the pilot selection process should place significant and measurable emphasis on selecting those candidates who have the attitude and social acumen to develop team unity and coordination.

References

French, J. R. P., & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp.150-167). Ann Arbor: University of Michigan, Institute for Social Research,

Green, P. E. & Vithala Rao. (1972). *Applied Multidimensional Scaling.* New York: Holt, Rinehart and Winston.

Hackman, J. R. (1993). "Teams, Leaders, and Organizations: New Directions for Crew-oriented Flight Training," In Wiener, E. L., Kanki, B. G., and Helmreich, R. L., (Eds.), *Cockpit Resource Management* (pp. 47–69). New York: Academic Press.

Helmreich, R. L. (1984). Cockpit management attitudes. *Human Factors, 26*(5), 583-589.

Helmreich, R. L., & Weilhelm, J. (1989). *Validating personality constructs for flightcrew selection: Status report on the NASA/UT Project.* NASA/UT Technical Memorandum, 893, Austin: University of Texas.

Kruskal, J. B. & Wish, M. (1978). *Multidimensional scaling.* Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills, Calif.: Sage.

Levine, J. H. (1979). Joint-Space analysis of ‘pick-any’ data: Analysis of choices from an unconstrained set of alternatives. *Psychometrika* 44 (March), 85-92.

Orlady, H.W., & Orlady, L.M. (1999). *Human factors in multi-crew flight operations*. Aldershot, England: Ashgate.

Schiffman, S. S., Reynolds, M. L. & Young, F. W.. (1981). *Introduction to Multidimensional Scaling.* New York: Academic Press.

StataCorp. (2005). Stata Statistical Software (Release 9.0) [Computer software]. College Station, TX: StataCorp.

International Journal of Applied Aviation Studies, Volume 6, Number 1

Book Review

Know the Risk: Learning from Errors and Accidents: Safety and Risk in Today's Technology by Romney Duffey and John Saull Published by Butterworth-Heinemann

Reviewed by
Frederick D. Hansen
Oklahoma State University
School of Educational Studies
700 N. Greenwood
Tulsa, OK 74106
U.S.A.
9180595-8406

Occasionally one stumbles across a book that offers a fresh perspective on an old problem. Researchers have been studying the problem of human error in aviation since Orville Wright crashed his airplane on September 17, 1908 killing his passenger Lieutenant Selfridge. In spite of new aircraft technologies, simulator training, Crew Resource Management, and the best efforts of researchers, the human contribution to accidents has remained consistently and frustratingly high.

In June 2003, U.S. Secretary of Defense Donald Rumsfeld challenged the military to cut the number of mishaps by fifty percent in two years. Is this a realistic goal? The Federal Aviation Administration set a goal to reduce fatal aviation accident rate by 80% by 2007. Is that a realistic goal?

Romney Duffey and John Saull offer a thesis that is both extremely persuasive yet simple and elegant that may offer a qualitative answer to the previous questions. By tapping extensive databases for aviation accidents, highway accidents, train accidents, nuclear power plant accidents, and even aircraft hijackings, the authors graphically plot the accumulated error reduction to produce what they claim to be a universal learning curve. In case after case, regardless of the industry analyzed, the resultant learning curve is similar. ". . . humans follow or obey a classic learning curve, which inexorably trends to lower error rates; and

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.

thus accident and error rate statistics follow a similar trend when accumulated experience is used as the basis for the measure of the activity”(p. 10).

Following the assumption that learning actually does occur and that we learn from our mistakes, the authors identify three error rates that are used throughout this book. The instantaneous rate (IR) or hazard rate is defined as “the rate at which errors occur in a fixed interval of time” (p. 16). The instantaneous rate can fluctuate greatly between time periods as we see with annual accident rates across many industries. The accumulated rate (AR) is then defined as “a measure of the rate at which errors are occurring over all experience, up to and including the latest interval” (p. 21). Finally, the authors define a constant rate (CR) to be a measure of the rate “if it is invariant with accumulated or increasing experience and falls to zero as infinite experience accumulates” (p. 23).

If the authors had stopped at this point, one would argue that at some point in time we could reduce our accident rate due to human error to zero – a highly desirable goal but one that has never been attained for a complex, technological system. The last component of the authors thesis considers the theory that errors can not be reduced below a certain minimum, regardless of how far we progress down the universal learning curve. The most interesting and informative component of this book revolves around their discussion and demonstration of the *minimum error rate* equation.

Whether you are prone to believe their data has demonstrated a cause and effect relationship between accidents and a learning curve based on accumulated errors or not, the authors provide an extensive range of evidence to support their claims. Using accumulated data for the sinking of ships, fatal airline crashes, general aviation accidents, nuclear power incidents, automobile traffic accidents, and medical errors, Duffey and Saull systematically demonstrate the apparent existence of a learning curve and more importantly a minimum error rate.

For risk managers, this book clearly suggests that effective risk management critically depends on the ability to track error reduction as it progresses down the universal learning curve. This ability naturally assumes that an accurate database exists that tracks errors and is timely. It also clearly suggests that over-ambitious goals of error reduction may produce some worthwhile gains in the instantaneous rate but that long-term gains will have to come from other sources – perhaps from a paradigm change in the man machine interface design.

In my opinion, this book should be required reading for safety practitioners, risk managers, and graduate students concerned with risk analysis and risk management. Know the Risk: Learning from Errors and Accidents: Safety and Risk in Today’s Technology will not be found in the library alongside books on Aviation Safety or Human Factors but it is worth the time to locate the book or better yet, go ahead and order one to keep in your own library.

CONTRIBUTOR INFORMATION

Content: The International Journal of Applied Aviation Studies publishes formal papers and training development articles, which supply the reader with alternative approaches to aviation challenges. The journal was born out of the concept of mutual support. Nations finding solutions to aviation problems at home could help other nations facing the same or similar problems in their country. Some of the many areas of interest of this journal are: operational support, regulatory standards, training support, airway facilities, air traffic control, airports and logistics, aviation security, and international training.

Manuscript: Rather than follow the traditional process of supplying three copies of your manuscript, we prefer that you send your manuscript via e-mail to the editor, L. Kay Chisholm (kay.chisholm@faa.gov). We prefer that each manuscript be prepared in Microsoft Word (at present, we are using Microsoft Office Professional Edition 2003). All figures and tables should be prepared as tif, .bmp, or .jpg files (preferably in 300 dpi), separately from the figure or table title. Tables may also be created in Microsoft Word. All figures and tables with figure and table titles should be incorporated into the manuscript. Please provide an abstract and a running header title for your manuscript. Additionally, please include your address, telephone number, and e-mail address after the names of the authors. If you want this information withheld during publication, please advise.

The manuscript should conform to the standards of the Publication Manual of the American Psychological Association (5th ed.). Be sure to reference this document when compiling the reference portion of your manuscript. We prefer italics rather than underlined titles or journal names.

Permissions: The editor assumes that all copied material has been included by permission of the original author or has been suitably cited.

Regulations: This journal publishes original work. We assume that you have not previously had this article published, nor submitted this article to another publication. The manuscript should be 35 pages or less, double-spaced, and conform to the standards of the Publication Manual of the American Psychological Association (5th edition).

CONSULTING EDITORS

Consulting editors are appointed by the Editor to represent all elements of aviation. An appointment recognizes professional and academic achievement. For consideration, submit a curriculum vita or statement of interest to Kay Chisholm, Editor, at kay.chisholm@faa.gov.