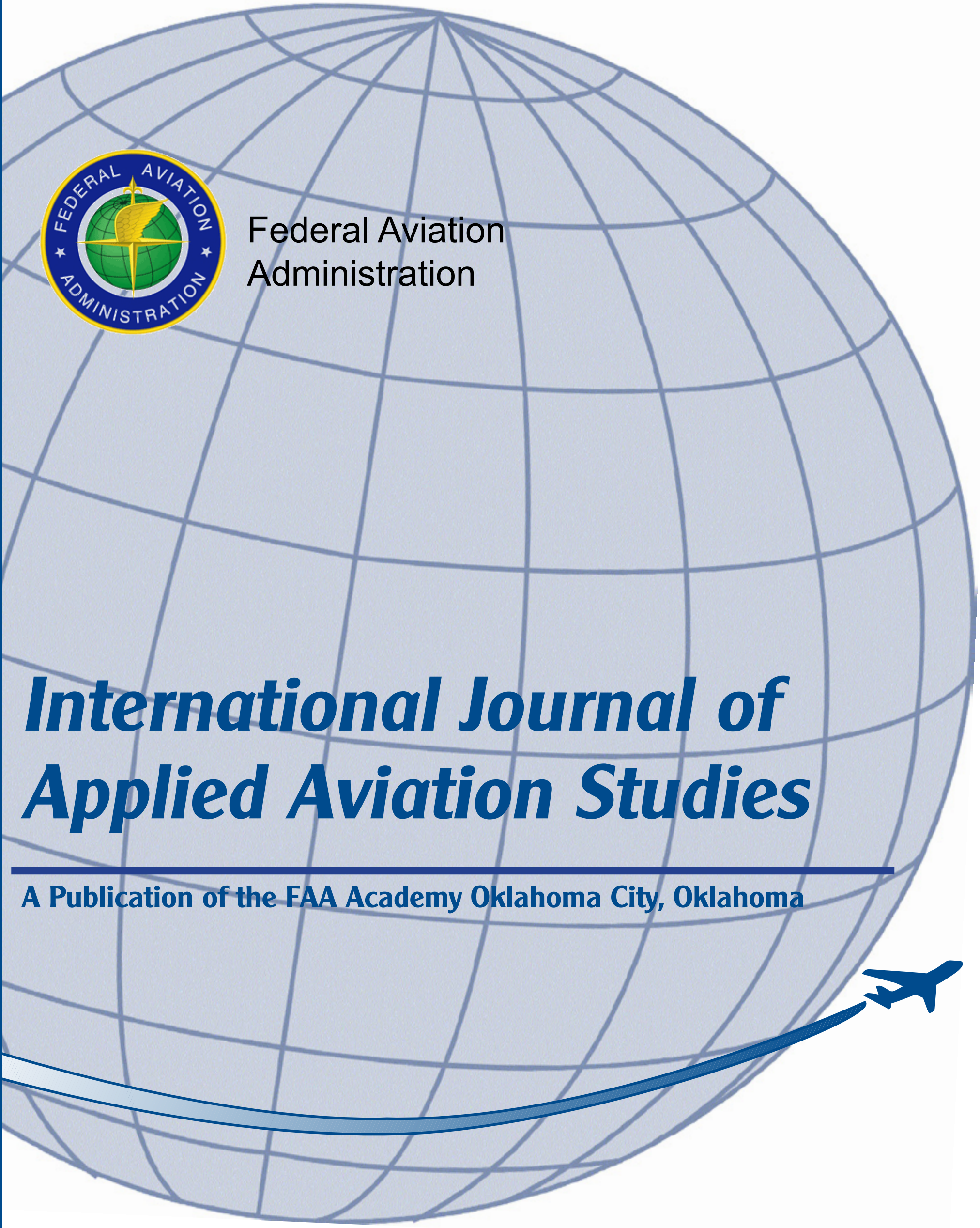
Federal Aviation
Administration



International Journal of Applied Aviation Studies

A Publication of the FAA Academy Oklahoma City, Oklahoma

Volume 7, Number 1, 2007
Government Printing Office

2007 *International Journal of Applied Aviation Studies* Volume 7, Number 1 A publication of the FAA Academy

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-530D
PO Box 25082
Oklahoma City, OK 73125

International Journal of Applied Aviation Studies
Volume 7, 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.

PUBLISHER

Academy Superintendent	Gary Condley,
Deputy Superintendent	Sunny Lee-Fanning(Acting)

PUBLICATION COMMITTEE

Manager, Regulatory Standards Division	David Long (Acting)
Manager, Airway Facilities Division	Jessie McMullen
Manager, Air Traffic Division	Patricia Garrambone
Manager, Airports & International Training Division	Doug Andresen(Acting)

EDITORIAL STAFF

Managing Editor	Kay Chisholm
Publication Editor	Beverly Laughead
Production Editor	Laurie Wilson
Associate Editor	Charlene Gahagan
Research Consultant	Todd P. Hubbard

CONSULTING EDITORS

International

Ruwantissa I.R. Abeyratne	ICAO, Montreal, Canada
Robert Bor	Royal Free Hospital, London, UK
Finian Connolly	Executive & Professional Training Institute, Drogheda, Co Louth, Ireland
Hugh David	Eurocontrol, East Sussex, UK
Sidney Dekker	Linköping Institute of Technology, Linköping, Sweden
Alois Farthofer	Human Factors Research, St. Georgen, Austria
Rodney Fewings	Cranfield University, Cranfield, Bedfordshire UK
Gerard J. Fogarty	University of Southern Queensland, Toowoomba, Australia
Jarle Gimmestad	Braathens Airlines, Fornebu, Norway
Don Harris	Cranfield University, Cranfield, Bedfordshire, UK
Irene Henley	University of Western Sidney, Sidney, Australia
Graham Hunt	Massey University, Auckland, New Zealand
Magnus Jacobssen	Linköping Institute of Technology, Linköping, Sweden
Lynn M. Jeffrey	Massey University Auckland, Auckland, New Zealand
K. Wolfgang Kallus	Karl Franzens University, Graz, Austria
Michael John Kay	RMIT University, Melbourne, Victoria, Australia
Elizabeth Mathews	ICAO, Montreal, Canada
Jeremy Mell	Ecole Nationale de l'Aviation Civile, Toulouse, France
Jan Noyes	University of Bristol, Bristol, UK
Esther Oprins	Air Traffic Control the Netherlands, Schiphol Airport, Netherlands
Jens Rolfsen	Det Norske Veritas, Høvik, Norway
Steve Thatcher	University of South Australia, Mawson Lakes, Australia
Claudia van de Wal	VALK Foundation, Netherlands
Lucas van Gerwen	VALK Foundation, Netherlands
Stephen Walsh	Interaction Training Associates, Netherlands
Rogier Woltjer	Linköping Institute of Technology, Linköping, Sweden
Beat Zimmermann	Air Navigation Institute, Thun, Switzerland

U.S.A.

Robert I. Aceves	St. Cloud State University, St. Cloud, MN
Vicki Ahlstrom	FAA, William J. Hughes Technical Center, Atlantic City, NJ
Amy Alexander	Aptima, Inc., Woburn, MA
R. Kurt Barnhart	Kansas State University, Salina, KS
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, Melbourne, FL
Raymond E. Cain, Jr.	Florida Memorial University, Miami Gardens, FL
Kim Cardosi	Volpe National Transportation Systems Center, Cambridge, MA
Stephen M. Casner	NASA Ames Research Center, Moffett Field, CA
Thomas R. Chidester	Civil Aerospace Medical Institute, Oklahoma City, OK
Robert H. Chisholm	Support Systems Associates, Melbourne, FL
David Conway	Southeastern Oklahoma State University, Durant, OK
Carmen Delgado-Morales	FAA Academy, Oklahoma City, OK
Archie Dillard	FAA National Headquarters, Washington, DC
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
Patricia Garrambone	FAA Academy, Oklahoma City, OK
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
Kevin High	Western Michigan University, Battle Creek, MI
Jerry Higley	FAA Academy, Oklahoma City, OK
Alan Hobbs	NASA Ames Research Center, Moffett Field, CA
Judy Holcomb	FAA Academy, Oklahoma City, OK
Amy Hoover	Central Washington University, Ellensburg, Washington
Todd P. Hubbard	Oklahoma State University, Stillwater, OK
Sig Illing	FAA Academy, Oklahoma City, OK
Kathi Ishimaru	FAA, Northwest Mountain Regional Office, Renton, WA
Jeffrey Johnson	St. Cloud State University, St. Cloud, MN
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, OK

Mary N. Kutz	Oklahoma State University, Stillwater, OK
Florian Jentsch	University of Central Florida, Orlando, FL
Victor LaSaxon	The Boeing Company, Midwest City, OK
Chien-tsung Lu	Central Missouri State University, Warrensburg, MO
Rebecca Lutte	University of Nebraska at Omaha, Omaha, NE
Richard L. Mangrum	Kent State University, Kent, OH
Carol Manning	Civil Aerospace Medical Institute, Oklahoma City, OK
Steven Marks	Oklahoma State University, Stillwater, OK
Royce Ann Martin	Bowling Green State University, Bowling Green, OH
Celeste Mayer	North Carolina State University, Raleigh, NC
Raymon M. McAdaragh	FAA, Hampton, Va
H. C. McClure	MAC Air Consulting, Queen Creek, AZ
Marjo Mitsutomi	University of Redlands, Redlands, CA
Kjell Mjøs	Norwegian University of Science and Technology, Oslo, Norway
Richard H. Mogford	NASA Ames Research Center, Moffett Field, CA
Kathleen Mosier	San Francisco State University, San Francisco, CA
Vahid Motevalli	The George Washington University, Ashburn, VA
Michael S. Nolan	Purdue University, West Lafayette, IN
Gary J. Northam	Parks College, St. Louis University, St. Louis, MO
Dale B. Oderman	Purdue University, West Lafayette, IN
Manoj Patankar	Parks College, St. Louis University, St. Louis, MO
Donald A. Petrin	Purdue University, West Lafayette, IN
Julia Pounds	Civil Aerospace Medical Institute, Oklahoma City, OK
Edward Pugacz	FAA, William J. Hughes Technical Center, Atlantic City, NJ
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
Robert Ripley	Auburn University, Auburn, AL
Charles L. Robertson	University of North Dakota, Grand Forks, ND
Nadine Sarter	The Ohio State University, Columbus, OH
Gregory L. Schwab	Indiana State University, Terre Haute, IN
J. Anthony Sharp	Ohio University Airport, Albany, OH
John W. Sheremeta, Jr.	American Airlines, Lake Ronkonkoma, NY
Mark A. Sherman	Ohio University, Athens, OH
Kenneth Sperry	The Boeing Company, Midwest City, OK
John E. Stewart II	U.S. Army Research Institute, Fort Rucker, AL
Alan J. Stolzer	Parks College, St. Louis University, St. Louis, MO
Judith B. Strother	Florida Institute of Technology, Melbourne, FL
Craig Gerald Stroup	American West Airlines, Phoenix, AZ
Judith Foss Van Zante	NASA Glenn Research Center, Cleveland, OH
Terry von Thaden	University of Illinois, Savoy, 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

Formal Papers

The human mind has a great capacity for remembering pictorial information. Our lead article, by *Arthur Estrada, Jennifer A. Keeley,* and *Patricia A. Leduc,* reports on a study, which tested the effectiveness of an innovative mnemonic strategy, the Intuitive Pictorial System (IPS). Analysis of user assessments and symbol recognition performance confirmed the value and advantages of employing the IPS to augment traditional methods. This information could initiate innovative training approaches and techniques for aviation.

System Safety requirements could become mandatory by 2010. This is an immediate challenge facing the aviation industry. Appling case-study philosophy and documentary analysis, *Chien-tsung Lus* and *Philip Bos'* article provides a comprehensive safety model, namely Aviation System Safety Management Model (ASSMM), combining Error Management, System Safety techniques and MIL-STD-882D to form a streamline risk analysis and accident prevention program for easier usage and safety reporting.

Responding to the call to provide scientific evidence that demonstrates the hypothesized relationships between the various components of safety culture and climate, *Erin E. Block, Edward J. Sabin,* and *Manoj S. Patankar's* paper presents a systems-level perspective for understanding and improving the safety climate in aviation. This study seeks to increase understanding of the ways in which various factors may influence safety outcomes for flight crewmembers.

William R. Knecht reports on the first experiment of a Nonveridical Aircraft Collision Avoidance System. This report describes the first major formal, successful test of 4CAS. 4CAS is a nonveridical, 4D aircraft collision avoidance system defined by three coordinate axes of aircraft heading, speed, and altitude. The results reported show that nonveridical displays may be useful in aircraft separation maintenance.

Adverse weather has a major impact on safety and operational efficiency. *Ulf Ahlstrom and Ferne Friedman-Berg's* study used simulation data to evaluate air traffic controllers' use of static and dynamic storm forecast displays that provided information about the direction of storm motion and future extrapolated positions. The study shows that dynamic storm forecast displays significantly reduced controller scan path areas, distances, and path durations compared to the static display. In addition, the study indicated a higher visual and cognitive workload during the static condition.

The quality assurance data to be analyzed by the web-based surveillance and auditing tool (WebSAT) is both qualitative and quantitative. Responses to checklist questions are quantitative. Open-ended responses, the second type of response for capturing maintenance errors, are qualitative. *Kunal Kapoor and Joel S. Greenstein's* research proposes to apply the statistical technique of multidimensional scaling (MDS) and the User Centered Design (UCD) method of Participatory Design (PD) to categorize open-ended responses into suitable performance metrics.

The focus of *William B. Rankin's* quantitative correlation study examines whether a relationship existed between the methods used for airport movement area driver training and the number of incursions at the Operational Evolution Plan (OEP-35) U.S. towered airports. The data from this study suggested a relationship exists between the methods used for airport driver training and the number of runway incursions at the largest U.S. towered airports.

Vahid Motevalli proposes that the Regional Oversight Organization (ROO) is the ultimate form of cooperation among civil aviation authorities. Motevalli's paper reviews different mechanisms of cooperation and reasons for regional oversight organizations becoming more attractive to Civil Aviation Authority (CAAs). It also examines critical issues associated with a framework for developing, sustaining, and making effective an oversight organization with a regional scope.

KC

International Journal of Applied Aviation Studies

Volume 7, Number 1 2007

Formal Papers

Developmental Reports

International Journal of Applied Aviation Studies, Volume 7, Number 1

Formal Papers

Facilitating Aviation Emergency Procedure Recall Using a Pictorial Mnemonic System

Arthur Estrada[1], Jennifer A. Keeley[2], and Patricia A. Leduc[1]

[1]*U.S. Army Aeromedical Research Laboratory*
Building 6901
Ft. Rucker, AL, U.S.A.

[2]*Touro University International*
5665 Plaza Drive, 3rd Floor
Cypress, CA 90630, U.S.A.

Abstract

Research shows providing students with memorization techniques improve their ability to recall information. Other studies show the human mind has a great capacity for remembering pictorial information. This study tested the effectiveness of a novel mnemonic strategy: the Intuitive Pictorial System (IPS). A pretest-posttest, control group design was applied to three groups based on experience levels. Descriptive and inferential statistics assessed the data to reveal any significant differences between the IPS and traditional training methods. Though the study's findings did not show the IPS produced performance gains superior to the traditional method, user assessments and symbol recognition performance demonstrated the utility and merit of the system as augmentation to traditional methods. The way the symbols were able to facilitate the recall of uncommon, unfamiliar terms and phrases in a naïve population to a level comparable to highly experienced pilots in one week highlighted the IPS's capacity to aid in the encoding of information into long-term memory. This information could lead to important innovations and be a valuable contribution to the U.S. Army and other military and civil aviation training.

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

Return to Table of Contents

Introduction

Ever since humans decided to take to the skies, they have had to deal with emergency situations created by the occasional malfunction or failure of aircraft components. The steps taken to ameliorate these conditions must be performed in an established manner and/or sequence. To this end, all student pilots must learn (memorize) and practice emergency procedures (EPs) established for specific emergency events for the particular model of aircraft in which they are being trained. The intent of this research was to examine the utility and merit of a novel system of intuitive symbols (the Intuitive Pictorial System or IPS) used as a mnemonic strategy in conjunction with the training of EPs to U.S. Army aviation students in order to facilitate the accurate recall of those procedures.

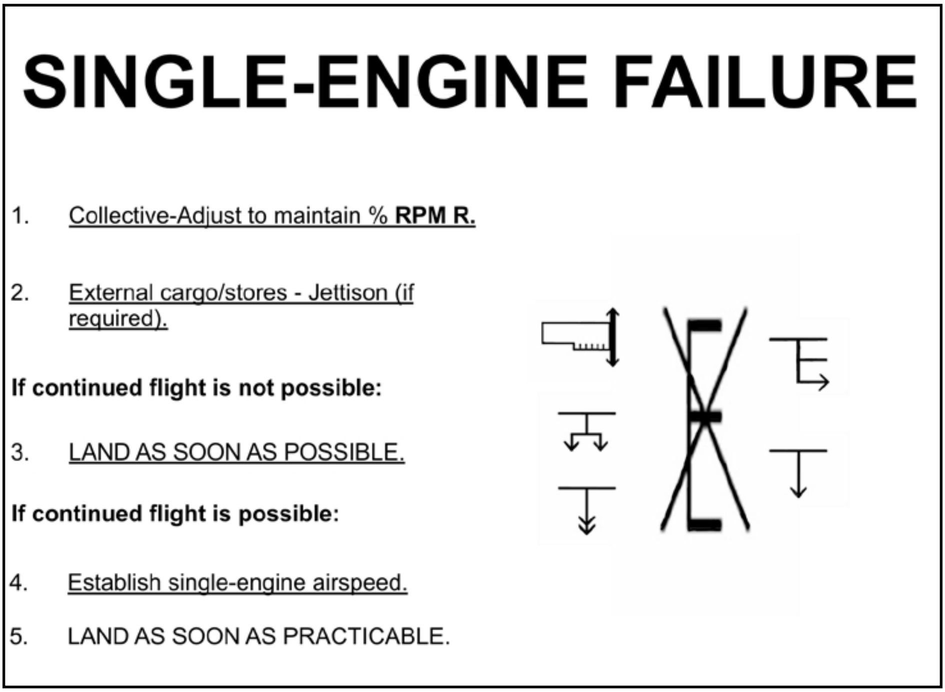
The U.S. Army's current practice of teaching aviation EPs requires students to learn the textual procedures through rote memorization. No standardized or formal guidance or techniques are provided. The U.S. Army is particularly demanding of their pilots in that it requires the memorization of an extensive number of EPs (many more than the other military services require), each of which usually contains multiple, purposefully-ordered steps. The satisfactory demonstration of EPs recall and performance is a requirement for the completion of flight school and the annual evaluations of pilots (Headquarters, Department of the Army, 1996).

The recall of the multitude of required EPs is a daunting effort. Actual aircraft emergencies are infrequent, and some are quite rare. Hence, without active recurring practice of the traditionally learned textual procedures, the ability of recall begins to decay significantly over a relatively short period. This premise is supported by the results of an anonymous survey of Army aviators (n = 194) conducted by the U.S. Army Aeromedical Research Laboratory (USAARL) that indicated that 16% of the respondents reported practicing their EPs daily, while 48% convey that they practice at least weekly (Estrada & Dumond, 2006). This reported need to practice so frequently, even after many years of experience, is indicative of a memory strategy that appears to be ineffective at transforming these procedural steps into reliably retrievable memory. Interestingly, 8% of the respondents were not satisfied with their own memorization methods, and 66% indicated an interest in learning new strategies to help retain their EPs.

There are a number of theories suggesting the mechanisms in which the brain transforms a perception (objects, text) into a meaningful perception (concepts) and then into a retrievable memory. Regardless of the theoretical mechanisms, the purpose of any learning is to transfer new information from working or short-term memory into long-term memory, memory that apparently has no capacity limits and holds information from minutes to an entire lifetime (Reed, 2004). Previous research has demonstrated that providing students with memorization techniques (mnemonic strategies) has resulted in improvements in their ability to recall learned information (Cox, 2001; Carney & Levin, 2003; Kleinheksel & Summy, 2003). Mnemonic strategies are systematic procedures for enhancing memory (Mastropieri & Scruggs, 1998) and are used to facilitate the acquisition of information because they assist in the memory encoding process, either by providing familiar connections or by creating new connections between to-be-remembered information and the learner's prior knowledge (Levin & Levin, 1990).

According to Bellezza (1992), memory experts learn to create mental pictures that endure in the mental space. Likewise, the IPS depicts each emergency situation and its procedural steps in a single pictorial form. The pictures that comprise the IPS are characterized as intuitive as they are formed with symbols representing aircraft parts and systems, and are accepted easily and are immediately recognizable to pilots. Thus, they require little cognitive effort in determining their meanings. For example, the capital E can be recognized as standing for the "engine." According to the rules of the IPS, an "X" over any symbol represents a failure of that component. Figure 1 presents the EP for a single-engine failure. Note that the large symbol represents the emergency situation (an E with an X over it); while the smaller symbols surrounding the larger one represent the various procedural steps. Hence, for memorization purposes, the entire symbol set is optimally remembered as a single picture

.



SINGLE-ENGINE FAILURE

1. Collective-Adjust to maintain % **RPM R.**
2. External cargo/stores - Jettison (if required).

If continued flight is not possible:

3. LAND AS SOON AS POSSIBLE.

If continued flight is possible:

4. Establish single-engine airspeed.
5. LAND AS SOON AS PRACTICABLE.

Figure 1. Example of IPS symbols application.

Theoretical Orientation

Learning and Cognition

Learning can be described as the gaining of knowledge, understanding, or skill with an outcome of a change in behavior because of the experience (U.S. Army Aviation Center, 2000). It is said to occur when the individual intentionally pays attention to the contents of working memory leading it to be absorbed into long-term memory (Herrmann, Raybeck, & Gruneberg, 2002). Reed (2004) defined cognition simply as the acquisition of knowledge; such acquisition involves many mental skills. Ashman and Conway (1997) elaborated that it involves taking in, storing, retrieving, transforming, and manipulating information. The process necessarily entails perception, awareness, judgment, the understanding of emotions, memory, and learning.

Too many times educators tell students what to learn yet fail to teach students how to learn. Knowing "how to learn" involves the learning of strategies. Strategies refer to the many methods in which we take in (encode), store, and retrieve (decode) information. Unfortunately, strategies used for enhancing learning are not an innate student ability. Cox (2001) observed that when left to their own devices, children would construct their own haphazard strategies. He noted that these variations in the effectiveness of these self-constructed strategies undoubtedly contribute to the wide individual differences in child development.

Theoretical frameworks for the brain's cognitive architecture have been presented and debated for many decades. Future research may discover the actual mechanisms of cognitive processing, but at present, allusions to theoretical concepts are the only recourse. In Cognition: Theory and Applications, Reed (2004) provided an account of the separate stages that researchers most commonly include in information-processing models. According to Reed, the sensory store provides brief storage for information in its sensory form. The perceptual information is then filtered which results in the blocking of some information, while other information is accepted or recognized. At the end of the filtering, all is lost unless a pattern is recognized (e.g., identifying a pattern as an animal or a written letter or word). Reed continued that the selection stage determines which information the person will try to remember. Following selection, the information is then moved into short-term (limited capacity in amount and duration) or long-term memory (with apparent unlimited capacities).

Learning Style Preferences

In a paper on approaches to teaching, Munro and Rice-Munro (2004) write,

> If a topic is important for students to learn, present it in a variety of ways that will stimulate learning…It's clear that there is no one instructional method that will reach all learners; therefore, it is up to those designing and delivering the instruction to offer a variety of approaches. (p. 28)

This notion that no one teaching method will reach all learners has emerged as the result of extensive research into individual student learning style preferences, specifically, their preferred sense modality of stimuli from which they most effectively take in, process, and store new information (Harrison, Andrews, & Saklofske, 2003). Simply stated, a learning style indicates an individual's preference for different types of information, the different ways in which it is perceived, and the rate at which the information is understood (Felder, 1993).

Visual Perception and Memory

Perceptions through the senses are the first steps in the cognitive process. Normal individuals acquire 75% of their knowledge through their sense of sight (U.S. Army Aviation Center, 2000). In the context of visual perception, Dretske (1995) made a distinction between sense perception and meaningful perception. Meaningful perception refers not to the objects one sees, but to how one perceives them. This meaningful perception embodies a judgment, belief, or recognition and requires conceptual skills, thus, requiring some level of visual cognition. Memory research in the 1960s by Haber (1970) suggested that there was one kind of memory for linguistic information (words, numbers, etc.) and another for pictorial information (scenes, pictures, etc.). Words are remembered, not as a picture of their letters, but as an idea or concept of the word(s). This requires several steps of encoding for proper storage and retrieval. Evidence from Haber's research indicated that this was not the case with pictorial information. He suggested that the pictorial image was received and stored permanently in its pictorial form. The review of this and other research regarding the cognitive quality of visual perception and visual memory is the foundation on which the Intuitive Pictorial System is based.

Research Hypotheses

Having established a theoretical and empirical basis for the study's construct, the study tested whether or not there was a difference in the demonstrated retention and recall of aviation EPs between the traditional and IPS teaching methods. Also examined was whether experience level or learning style preference had an effect on learning. Another analysis considered the manner in which the IPS was subjectively assessed. That is, whether or not experience level and learning style preference had any effect on their assessments. The ultimate goal of the study was to determine whether the IPS was a useful pictorial mnemonic for remembering textual aviation EPs. To quantify this, data were examined to ascertain whether there was a difference in the ability to recognize and remember the IPS symbols from Day 1 to Day 7.

Methods

Research Design

The study protocol was approved in advance by the U.S. Army Aeromedical Research Laboratory Human Use Committee. Each subject provided written informed consent before participating. Primarily, this study tested the merits of the IPS through a pretest-posttest control group design applied to three distinct groups based on their flight experience levels and, thus, their exposures to the traditional training method. The first group was composed of instructor pilots considered highly experienced in traditional EP training. The second group, student pilots with minimal experience, was undergoing training in the U.S. Army's UH-60 Black Hawk helicopter qualification course. The third group was composed of students waiting to start flight school and thus, naïve to EP training.

The three experience groups were each divided by random assignment to receive either the IPS or traditional training methods. On Day 1, the knowledge of those assigned to the IPS group were pretested (Emergency Procedures Knowledge Test), trained in the use of the novel IPS, and then post-tested with the same test. Similarly on Day 1, those assigned to the traditional group were pretested, received training/review of the EPs in the traditional manner, and then post-tested. The IPS and traditionally trained groups were then post-tested seven days later. Each posttest was intended to measure the learning, retention, and recall of the IPS trained group compared to that of the equivalent traditionally trained group.

In order to test the IPS's merit as an effective learning strategy and the intuitiveness and memorable qualities of the IPS symbols themselves, the IPS trained groups received two additional items: the Subjective Assessment Survey and the Symbol Recognition Test. These items were administered immediately following training of the IPS on Day 1 and then again Day 7.

Participants and Setting

As described above, the study consisted of three groups based on their experience levels. All participants were recruited from personnel assigned to Fort Rucker, Alabama, through local advertisement and solicitation. All training and testing for this study occurred in classrooms at Lowe Army Heliport or in a USAARL meeting room.

Sample Size

The sample size required to achieve a sufficient power was calculated specifying a desired power of .80, and a level of .05, and a large effect size (d = 1.00). The effect size of 1.00 was based on the results of a study by Carney and Levin (2003). Carney and Levin's work examined the effects of training undergraduate students in the use of mnemonic strategies (pictorial representations) on their memory and recall of unfamiliar hierarchical information, which was sufficiently similar to the objectives of this study. The power calculation indicated 30 participants per group (15 per IPS and 15 per traditional) were required to ensure adequate power. Hence, this study required a total size of 90 participants, 30 per experience level group.

Data Collection Tools

The Emergency Procedures Knowledge Test for all groups consisted of a 15-question fill-in-the-blank written examination covering 17 selected EPs listed in the UH-60 Black Hawk Operator's Manual (Headquarters, Department of the Army, 2003). The test had 25 answers with each answer worth four points; the final test scores ranged from 0% to 100% correct.

The VARK© Learning Styles Questionnaire (VARK, n.d.) provided a profile of user learning preferences: visual, aural, read/write, kinesthetic, and multimodal. (Permission to use the VARK questionnaire was granted by Mr. Neil D. Fleming, Christchurch, New Zealand.)

The Subjective Assessment Survey consisted of a series of questions aimed at determining the merits of the IPS as a mnemonic strategy based on the opinions of the users. The responses were made on a five-point Likert scale ranging from strongly disagree to strongly agree. Solicitation for additional comments was made.

The Symbol Recognition Test was a 20-symbol, fill-in-the-blank test that was intended to gauge the intuitiveness and memorable quality of the IPS symbols. The test had 20 answers with each answer worth 5 points; the final test scores ranged from 0% to 100% correct.

Testing procedures

All identifying data were separated from test data and were stored in a locked safe in the USAARL where they will remain for a minimum of three years following completion of the study. Each potential participant was briefed on the objectives of the study. If continued participation was indicated, each participant was given adequate time to review and understand all the information in the informed consent form. After signing the informed consent, the participants were randomly assigned to either an IPS or traditionally trained group and assigned to the appropriate experience level group (30 participants per group). Participants were instructed when to report for Day 1. Each group was managed on different days and times of the week in order to eliminate any interaction between groups. In order to avoid any subject bias possibly introduced by the words “novel” and “traditional,” the terms “Method 1” and “Method 2”, respectively, were used whenever interacting with the participants.

On Day 1, each participant completed a VARK© Learning Preference Questionnaire. Following that, each group received a standardized general information briefing and was administered the Emergency Procedures Knowledge Test as a pretest. Next, depending on their group assignment, members were provided training of 17 selected UH-60 EPs by either the IPS or traditional method.

At the completion of the training, the Emergency Procedures Knowledge Test was administered again as a measure of learning. In addition, members receiving the IPS method were administered the Symbol Recognition Test and were provided with a paper copy of the Intuitive Pictorial System Symbols and Rules, for additional review and study. Members of all groups received a paper copy of their appropriate PowerPoint presentation slides. On Day 7, the same Emergency Procedures Knowledge Test was administered as a second posttest. Following the posttest, a Subjective Assessment Survey was administered to the IPS trained participants in order to gather opinion-based data regarding the merits of the novel IPS. The Symbol Recognition Test was re-administered to the IPS trained groups.

Data analysis approach

All statistical analyses were conducted using SPSS® 12.0 with statistical significance set at an alpha level of .05. The independent variables were the methods of training, experience levels, and learning style preferences. The dependent variables were derived from the scores of the tests described above.

Results

To answer whether the employment of the IPS would demonstrate improved recall of EPs over that of the traditional teaching method in the sample, two comparisons were made of the Emergency Procedures Knowledge Test change scores: 1) Day 1 Pretest and Day 1 Posttest, and 2) Day 1 Posttest and Day 7 Posttest. During the initial data exploration, a Shapiro-Wilk test of normality was performed that showed that these data were not normally distributed [$W(45) = .913$, $p = .003$ and $W(48) = .897$, $p = .001$, respectively]. Attempts to transform the data did not change this condition. Therefore, the nonparametric Mann-Whitney U test was used in lieu of the originally planned independent-samples t test as the test for comparing the performance of the IPS and traditional instructional methods.

In the first comparison, the traditional method demonstrated a significantly greater average improvement ($U = 734.50$, $p = .008$) in performance than did the IPS method (19.2 vs. 8.75, respectively). This finding would indicate that the IPS did not improve, and could actually have impaired, recall. A possible explanation for this result is presented in the discussion section later. A second comparison, showed that there were no significant differences in the performance changes between method groups from Day 1 to Day 7 ($U = 1051.00$, $p = .823$).

A Wilcoxon's Signed-Rank test (alternative to paired samples t test) revealed that each group made statistically significant improvements in their posttest performances from Day 1 to Day 7. The IPS trained group improved from a mean of 65.00 to 72.83 ($p = .007$), while the traditionally trained group improved from 66.49 to 73.51 ($p = .007$).

To test whether or not there was a learning difference with respect to experience levels, a Kruskal-Wallis test was employed, followed by a Median test used to determine whether any patterns represented significant group differences. In an analysis of these data, the Kruskal-Wallis test detected that significant differences existed in the changes in the Emergency Procedures Knowledge Test scores from Day 1 Pretest to Day 1Posttest [IPS method: $\chi 2(2, N = 48) = 24.442$, $p < .001$; traditional method: $\chi 2(2, N = 45) = 22.905$, $p < .001$] and from Day 1 Posttest to Day 7 Posttest [IPS method: $\chi 2(2, N = 48) = 17.012$, $p < .001$; traditional method: $\chi 2(2, N = 45) = 6.155$, $p = .046$].

The Median test indicated that both the IPS and traditionally trained groups of the naïve group achieved significant improvements in their scores from the Day 1 Pretest to the Day 1 Posttest [$\chi 2(2, N = 48) = 20.761$, $p < .001$ and $\chi 2(2, N = 45) = 19.821$, $p < .001$, respectively]. The Median test also revealed a significant difference by experience levels in performance gains (change scores) between the Day 1 Posttest and the Day 7 Posttest [IPS method: $\chi 2(2, N = 48) = 13.746$, $p = .001$; traditional method: $\chi 2(2, N = 45) = 8.820$, $p = .012$]. The highly experienced group did not show performance gains greater than the median of their respective groups. This is likely due to a ceiling effect. That is, their preexisting high level of EP knowledge competence may be so high that treatment effects are not perceptible or are obscured.

The Kruskal-Wallis and Median tests were again employed to test whether or not there was a difference in recall performance between the two methods with respect to individual learning style preferences. As a result of the VARK questionnaire, the majority (73%) was classified as multimodal learners, preferring a variety of presentation modalities. The remaining population was made up of 3% aural, 9% reading, and 15% kinesthetic learners. Note that no participants were categorized as strictly visual learners. The results of the Kruskal-Wallis and Median tests showed that individual learning style preferences had no statistically significant effect on performance improvements. In addition, nonparametric correlations (Spearman's rho) were performed. These results revealed no significant relationships either.

Nonparametric correlation procedures were performed to measure the relationship of recall performance (as measured by the scores of the Day 1 and Day 7 Posttests and the change scores from Day 1 to Day 7) and of the IPS trained groups' subjective assessments of the IPS (from positive to negative). In other words, did the IPS trained groups' performance and performance changes have an influence on their opinions of the IPS? The results indicated that no statistically significant correlations existed.

To determine if individual learning style preferences had any effect on aviators' subjective assessment of the IPS, a Pearson's Chi-square test was employed. The results suggest that the two variables were independent of each other, and hence learning style preferences had no statistically significant effect on the subjective assessment of the IPS.

A Pearson's Chi-square test was used to determine whether experience levels were independent of the subjective assessment of the IPS based on the categories derived from the Subjective Assessment Survey scores. The results showed that the two variables were associated [χ 2(4, N = 48) = 20.657, p < .001]. An examination of Table 1 shows that the degree of positive assessment is inversely related to the extent of aviation experience. As such, the results indicate that levels of experience are related to the subjective assessment of the IPS.

Table 1

Subjective Assessments

	Positive	Positive to Neutral	Neutral	Neutral to Negative	Negative	Totals
Experienced Pilots	1	12	0	2	0	15
Student Pilots	4	13	0	0	0	17
Naïve Students	12	4	0	0	0	16

Even though significant differences were found in IPS assessments depending on aviator experience, the overwhelming majority, nearly 96% of the IPS trained population, rated the IPS in a positive manner (above neutral) (Table 2).

Table 2
Frequency of Overall IPS Rating.

	N	Percent
Positive	17	35.4
Positive to Neutral	29	60.4
Neutral	0	
Neutral to Negative	2	4.2
Negative	0	
Total	48	100.0

Finally, the Subjective Assessment Survey solicited additional comments regarding the IPS, the intuitiveness of the symbols, and recommendations for future applications. A review of the remarks indicated that 50% (16) were laudatory in nature; 25% (8) recommended specific symbol changes; 15.6% (5) were laudatory of the IPS, yet negatively critical of some of the symbols; and 9.4% (3) were negatively critical of the IPS as a mnemonic system, citing the possibility of causing confusion and/or misinterpretation.

A paired samples *t* test was used to determine whether there was any difference in the ability to recognize and remember the IPS symbols from Day 1 to Day 7. The Day 1 test was administered immediately following the participants' introduction to the IPS symbols, and even with this limited exposure to the symbols (less than 30 minutes) prior to the test, each group scored, on average, at least 75% correctly (Table III), with 22.9% achieving a perfect score.

Table 3
Symbol Recognition Score Means by Experience Level

	Day 1 Symbol Recognition Test	Day 7 Symbol Recognition Test
Experienced Pilots	85	85
Student Pilots	92	81
Naïve Students	75	85

Overall, the participants' symbol recognition scores did not differ significantly [*t*(47) = .290, *p* = .773] from one week (*M* = 84.38) to the next (*M* = 83.65). The highly experienced group Day 1 and Day 7 test scores (*M* = 85.00 and *M* = 85.00, respectively) were identical.

In contrast, the minimally experienced and naïve groups showed significant differences in their test performances from Day 1 to Day 7, albeit in different directions. The minimally experienced group showed a statistically significant decline in performance from one week (*M* = 92.35) to the next (*M* = 81.18) [*t*(16) = 3.297, *p* = .005], while the naïve group demonstrated a statistically significant [*t*(15) = -3.183, *p* = .006] improvement in their test performance from Day 1 to Day 7 (from *M* = 75.31 to *M* = 85.00).

The minimally-experienced group showed a decline from an average score on the first test that was seven points *higher* than the highly-experienced group (92% vs. 85%, respectively) to a score one week later that was just four points *lower* the highly-experienced group (81% vs. 85%). Also, note that the naïve group achieved the same level of symbol recognition on Day 7 as the highly-experienced group (both with 85% scores). The importance of these observations is that on both tests, the average of the least experienced participants scored nearly the same or better than the highly experienced group, who have a minimum of four years experience and familiarity memorizing the textual EPs. The performance of the less experienced groups, especially the naïve group, in closely matching the highly experienced group demonstrates the ability of the symbols to bring back uncommon text for which there was minimal or no familiarization.

Discussion

The results indicated that the IPS did not improve retention and recall overall as compared to the traditional method. In light of previous research extolling the ease and capacity for remembering pictorial representations (Haber, 1970; Potter, 1976; Hollingworth & Henderson, 2002), this finding was unexpected. Moreover, the overall traditionally trained group achieved a statistically significant improvement in their scores. The large disparity between group performances was initially puzzling; however, a plausible explanation for the disparity may be the difference in the amount of information presented to each group. It is important to note that the Day 1 Emergency Procedures Knowledge Posttest was administered immediately following the EP Instructional PowerPoint Presentation and that two-thirds of the participants had previous memorization experience with the EPs. Thus, the Emergency Procedures Instructional PowerPoint Presentation served the majority of the traditionally trained group as a review and reinforcement of existing/established memory strategies. In contrast, those in the IPS trained group were presented with the same EPs and were directed to employ a new memorization strategy as their memorization technique. The 30-minutes to one-hour interval between Day 1 Emergency Procedures Knowledge Pretest and Posttest may not have been sufficient time to assimilate the amount of new information presented.

The plausibility of this explanation is supported when the Day 1 and Day 7 Posttests are compared, and the results are inconsistent with the findings above. In this comparison, the average score improvements made by the IPS and traditionally trained groups did not differ significantly from each other. With regard to the research question, the IPS still did not improve retention and recall over that of the traditional method; however, neither method was significantly superior to the other. Moreover, participants receiving either method of instruction made statistically significant improvements, on average, in their Emergency Procedures Knowledge Posttest performances from Day 1 to Day 7.

In determining if there was a difference in recall performance between the IPS and traditional methods with regard to experience levels, the statistics revealed that significant differences in performance were influenced by experience levels with the only difference appearing among the minimally experienced group. Neither training method had an affect on the performance of the highly experienced group who showed no significant improvements in any of their tests. This may have been due to a ceiling effect. In other words, participants of this group started with relatively high pretest scores, which then remained high on their two posttests.

As in the case with the highly experienced group, the training method made no significant difference in the performance improvements of the naïve group, as both IPS and traditionally trained groups made significant gains in their performances. The difference between Day 1 and Day 7 Posttest performances of the IPS and traditionally trained groups in this naïve population may be due to the difference in the amount of new information presented to each group on Day 1.

A significant difference between instructional methods was detected in the minimally experienced group. The Median test for the minimally experienced group illustrates no practical differences between IPS and traditionally trained groups for the performance improvements from Day 1 Pretest to Day 1 Posttest but does show differences in performances from Day 1 to Day 7. Furthermore, it was the traditionally trained group that performed better (an 11 point average improvement), while the IPS group's average score decreased by 3 points. These findings seem to favor the superiority of the traditional method over the IPS with regard to minimally experienced aviators. However, the findings are likely confounded by the potential effects of the inability to control the time participants dedicated to studying between Day 1 and Day 7 tests. Because the sample population was actively engaged in some capacity in the U.S. Army's flight school program as instructors, active students, or inactive students, and not subjected to a completely controlled laboratory environment, study time could not be regulated or controlled.

The VARK Questionnaire was employed to classify participants' learning style preferences. According to theory, individuals differ in the manner from which they best absorb information (Cassidy & Eachus, 2000; Harrison, Andrews, & Saklofske, 2003) based largely on the preferred learning modality (Zapalska & Dabb, 2002). Specifically, those with style preferences should show the greatest ability to recall material presented in that manner (Krätzig & Arbuthnott, 2006). This study was composed of very few members having a single learning style preference with the majority (68 of 93) classified as having multimodal learning preferences. Despite reports praising the virtues of matching learning style preference with modality of presentation (Lohri-Posey, 2003; Munro & Rice-Munro, 2004), this study's findings are consistent with those of Loo (2004) and Krätzig and Arbuthnot, who report weak to no significant link between learning style preference and objective memory performance. In any case, the findings, at least in this sample, indicate that neither the IPS nor the traditional training method is prejudicial toward any single or multimodal learning style preferences.

According to the Instructor Pilot Handbook (U.S. Army Aviation Center, 2000), the *Principle of Effect* states that learning is strengthened when accompanied by a pleasant or satisfying feeling. Conversely, learning experiences that produce feelings of frustration, confusion, and futility weaken learning. To achieve a sense for the emotional effect of the IPS on the treated membership groups, participants' performances and performance gains were correlated to the manner in which they assessed the IPS. Since no statistically significant correlations were found between recall performance and the subjective assessment of the IPS, no inferences or presumptions of training effect, either positive or negative, can be made based upon the relationship of these two variables.

An inquiry into the possibility that learning style preference might predispose one's subjective assessment of the IPS was performed. Stated differently, the goal was to determine if those more likely to prefer learning in other than visual ways might have a less positive opinion of the IPS than visually oriented learners might. The analysis showed that the two variables (learning style preference and subjective assessment categories) were independent and unrelated and, operationally, that learning style preference has no affect on one's opinion of the IPS as a mnemonic strategy.

The results indicated that experience does have an affect on the subjective assessment of the IPS. Simply put, the more experience, the less favorable the assessment. That said, the IPS received only two assessments that were negatively inclined, both by members of the highly experienced group. Even though all other assessments were slanted in a positive direction, the significant relationship between experience and subjective assessment was evident.

A reasonable explanation for the significant relationship may be based on a common resistance towards change. According to the expectancy theory (Lines, 2004), "people consciously choose courses of action, based upon perceptions, attitudes, and beliefs, as a consequence of their desires to enhance pleasure and avoid pain" (p. 198). Expectancy theory predicts that resistance will occur if 1) the individual has expectancies that the relationship between a change in behavior and performance is uncertain, 2) the link between performance and outcome is uncertain, and 3) the outcomes have negative value to the individual (Hope & Pate, 1988).

Experience in aviation implies learned and established knowledge and skills. Once established, any alterations or redefinitions naturally lead to some people feeling uncomfortable or threatened (Walley, 1995). Aviators, being professional people, value their professionalism and hard-won skills. Any changes that threaten to make these skills obsolete are likely to encounter resistance, especially with a less technically competent aviator who may become defensive to preserve a competent self-image (Jensen, 1998). In addition, unfamiliar novel systems, such as the IPS, provoke barriers to change, such as fear of the unknown, low trust, arrogant attitudes, and resistant organizational cultures (Kane & Darling, 2002). In light of the preceding discussion, the discovery that experience levels affect the assessment of the IPS is logical and expected.

The Subjective Assessment Surveys answered whether the IPS was appraised as a useful mnemonic for the intended purpose in a straightforward and unproblematic manner. Ninety-six percent of the sample population rated the system in a favorable manner. Support for incorporating the IPS into flight training was evidenced by noting that only one of the participants would not recommend the system for use in the U.S. Army. No respondent felt that the IPS was too complex for new flight students, and none recommended that it not be taught to future flight students. These findings support previous research (Haber, 1970; Cherry, Dokey, Reese, & Brigman, 2003) that demonstrated peoples' predilection for pictorial representations and their usefulness as aids to memory.

One goal of this research question was to determine if the IPS symbols were easily recognized (intuitive) and, thus, had memorable qualities. Since the first Symbol Recognition test was administered within 30 minutes of the participants ever seeing the symbols, a mean score of 84.38 for the sample population was impressive and indicative of the symbols' intuitive nature and memorability. The fact that every experience group averaged high scores immediately following their initial presentation showed the symbols to be intuitive regardless of prior knowledge and familiarity with the represented terms. These findings may be the single most important indication of the merits and utility of the IPS as an aid to memory.

When the Symbol Recognition Test was administered after one week, the sample aggregate mean score had not changed significantly. When each experience group was examined individually, the highly experienced group's mean performance showed no change. However, the naïve group showed a significant performance improvement over the one-week period, while the minimally experienced group showed a decline in performance. Although a significant drop in the average score by the minimally experienced group, the lower score still demonstrates strong and successful recall of the symbols nearly equivalent with that of the highly experienced group. The test results and comparisons suggest strongly that the symbols are indeed intuitive and memorable.

Conclusions

Although the findings provide substantial evidence that the IPS did benefit its users, the conditions of the study posed some challenges. For one, aviators are professionals who are proud of their knowledge and skills. Hence, it was apparent during the recruitment of participants that there was a general aversion to tests that might challenge or question their knowledge. There were palpable differences between experience groups and individuals in their motivation and their commitment to the research effort. In addition, limitations in time and funding dictated the measurement of performance gains of at least ninety participants over a one-week period. Since memorization strategies such as the IPS are used to enhance long-term memory, longitudinal studies over much longer periods may provide a much clearer, substantive depiction of the effectiveness of the IPS compared to that of the tradition method. In addition, the value of the IPS during emergency situations in simulated or actual flight remains to be tested.

Although the study's findings did not show that the IPS produced performance gains superior to those of the traditional method over the one-week trial, the findings did demonstrate the utility and merit of the system as an augmentation to traditional textual procedures. The positive results of the Subjective Assessment Surveys and Symbol Recognition Tests indicated the general acceptance, intuitiveness, and memorability of the prototype symbols. The results indicate that the IPS is most beneficial to a naïve population. The manner in which the symbols were able to facilitate the recall of uncommon, unfamiliar terms and phrases in a naïve population to a level comparable to that of highly experienced pilots in just one week, highlights the potential for such a mnemonic strategy to aid in the encoding of information into long-term memory. The promise in incorporating a refined IPS into the U.S. Army flight educational program would be in reducing the expense and time it takes to teach, learn, and maintain complex aviation EPs, thus enhancing aviation safety and preserving vital resources. A standardized pictorial system of aviation procedures may not just benefit the recall performance of individual graduate and student pilots but ultimately could be a valuable contribution to aviation safety by providing a useful, abbreviated presentation for reference during actual emergency situations.

Acknowledgements

The authors would like to acknowledge LTC Robert M. Wildzunas, Ph.D., and Dr. John S. Crowley for their support and guidance during this project.

Disclaimer

References

Ashman, A.F. & Conway R.N.F. (1997). *An Introduction to cognitive education. Theory and applications.* New York, NY: Routledge.

Bellezza, F.S. (1992). The mind's eye in expert memorizers' descriptions of remembering. *Metaphor and Symbolic Activity, 7*(3 & 4), 119-133.

Carney, R.N. & Levin, J.R. (2003). Promoting higher-order learning benefits by building lower-order mnemonic connections. *Applied Cognitive Psychology, 17*(5), 563-575.

Cassidy, S. & Eachus, P. (2000). Learning styles: An overview of theories, models, and measures. *Educational Psychology, 24*(4), 419-444.

Cherry, K.E., Dokey, D.K., Reese, C.M. & Brigman, S. (2003). Pictorial illustrations enhance memory for sentences in younger and older adults. *Experimental Aging Research, 29*(3), 353-370.

Cox, B.D. (2001). Children's use of mnemonic strategies: Variability in response to metamemory training. *The Journal of Genetic Psychology, 155*(4), 423-442.

Dretske, F. (1995). Attention and visual perception: Analyzing divided attention. In D.N. Osherson (General Ed.) & S.M. Kosslyn & D.N. Osherson (Vol. Eds.), *An invitation to cognitive science: Vol. 2. Visual Cognition (2nd ed.)*. Cambridge, MA: MIT Press.

Estrada, A. & Dumond, S. (2006). *A survey of aviator perceptions of aviation emergency procedure training and recall* (Report No. 2006-06). Fort Rucker, AL: United States Army Aeromedical Research Laboratory.

Felder, R. (1993). Reaching the second tier: Learning and teaching styles in college science education. *Journal of College Science Teaching, 23*(5), 286-290.

Haber, R.N. (1970). How we remember what we see*. Scientific American, 222*(5), 104-112.

Harrison, G., Andrews, J., & Saklofske, D. (2003). Current perspectives on cognitive learning styles. *Education Canada, 43*(2), 44-47.

Headquarters, Department of the Army. (1996). *Aircrew training program commander's guide to individual and crew standardization.* (Training Circular 1-210, with Change 1). Washington, DC: U.S. Government Printing Office.

Headquarters, Department of the Army. (2003). *Operator's manual for UH-60A helicopter, UH-60L helicopter, EH-60A helicopter*. (Technical Manual 1-1520-237-10, with Change 1). Washington, DC: U.S. Government Printing Office.

Herrmann, D., Raybeck, D. & Gruneberg, M. (2002). *Improving memory and study skills. Advances in theory and practice*. Seattle, WA: Hogrefe & Huber Publishers.

Hollingworth, A. & Henderson, J.M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113-136.

Hope, J.W. & Pate, L.E. (1988). A cognitive expectancy analysis of compliance decisions. *Human Relations, 41*(10), 739-751.

Jensen, R.S. (Ed.). (1998). *Aviation Psychology*. Brookfield, VT: Gower Publishing Company.

Kane, P.R. & Darling, L.A.H. (2002). Learning the pedagogy of change. *Independent School, 62*(1).

Kleinheksel, K.A. & Summy, S.E. (2003). Enhancing student learning and social behavior through mnemonic strategies. *TEACHING Exceptional Children,36*(2), 30-35.

Krätzig, G.P. & Arbuthnott, K.D. (2006). Perceptual learning style and learning proficiency: A test of the hypothesis. *Journal of Educational Psychology, 98*(1), 238-246.

Levin, M.E. & Levin, J.R. (1990). Scientific mnemonomies: Methods for maximizing more than memory. *American Educational Research Journal, 27*(2), 301-321.

Lines, R. (2004). Influence of participation in strategic change: resistance, organizational commitment and change goal achievement. *Journal of Change Management, 4*(3), 193-215.

Lohri-Posey, B. (2003). Determining learning style preferences of students. *Nurse Educator, 28*(2), 54.

Loo, R. (2004). Kolb's learning styles and learning preferences: Is there a linkage? *Educational Psychology, 24*(1), 99-108.

Mastropieri, M.A. & Scruggs, T.E. (1998). Enhancing school success with mnemonic strategies. *Intervention in School & Clinic, 33*(4).

Munro, R.A. & Rice-Munro, E.J. (2004). Learning styles, teaching approaches, and technology. *The Journal for Quality and Participation, 27*(1), 26-32.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509-522.

Reed, S.K. (2004). *Cognition. Theory and Applications* (6th ed.). Belmont, CA: Wadsworth/ Thomson Learning.

U.S. Army Aviation Center. (2000). *Instructor Pilot Handbook, Instructing Fundamentals* (PFN 0680-4). Fort Rucker, AL.

VARK (n.d.). *A Guide to Learning Styles*. Retrieved June 16, 2005, from http://www.vark-learn.com /english/page.asp?p=copyright.

Walley, C. (1995). Looking at school change. *Childhood Education, 71*(5), 258-260.

Zapalska, A.M., & Dabb, H. (2002). Learning Styles. *Journal of Teaching in International Business, 13*(3/4), 77-97.

International Journal of Applied Aviation Studies, Volume 7, Number 1

System Safety Application: Constructing a Comprehensive Aviation System Safety Management Model (ASSMM)

Chien-tsung Lu

Philip Bos

Department of Aviation, TRG 210
University of Central Missouri
Warrensburg, MO 64093
660-543-8092
ctlu@ucmo.edu
bos@ucmo.edu

&

William Caldwell, Ph.D.
Southern Illinois University Carbondale
Carbondale, IL 62901
618-453-8898
caldwel@siu.edu

Abstract

System Safety concepts have been developed and presented to aerospace engineers since the early 1960s. Yet to date, most air carriers and FBOs have not implemented system safety processes. The lack of implementation is probably because the value of System Safety is still unclear to them (USGAO, 2005a). Nevertheless, the immediate challenge facing the aviation industry is the implementation of System Safety requirements, which could become mandatory by 2010 (USGAO, 2005b). While applying System Safety methods remains optional, academia possesses an opportunity to help explain its essentiality and demonstrate its usefulness. This paper uses a hybrid of case study philosophy and documentary analysis to initiate research objectives, which include: 1) briefly reviewing the FAA's voluntary safety programs and revealing operational problems, 2) introducing the essential techniques and tools of System Safety explaining the complexity of accident causality, 3) proposing and demonstrating the process of a comprehensive System Safety Management Model, and 4) suggesting future research topics for airlines, airports, aviation scholars, and government.

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

Return to Table of Contents

Introduction

Historically, safety has been the mission priority and universal norm for the worldwide aviation industry. The priority of safety is especially prevalent in the airline sector. The 9/11 terrorist attacks in 2001 provided the impetus for further air transportation safety and security measures. Safety and security has become the utmost importance and, to a great extent, has triggered numerous studies and research involving safety and security performance. In the official 9/11 Commission Report, a multi-layer redundant system is recommended to effectively secure needed safety quality and security levels (National Commission on Terrorist Attacks, 2004). As addressed in the Report:

> The FAA set and enforced aviation security rules, which airlines and airports were required to implement. The rules were supposed to produce a "layered" system of defense. This means that the failure of any one layer of security would not be fatal, because additional layers would provide backup security. (National Commission on Terrorist Attacks, 2004, p. 81)

The FAA's Safety Management Program

The idea of redundancy as described in the aforementioned report has created a critical challenge to aviation safety. In fact, in the FAA, the Office of System Safety is empowered to lead Aviation System Safety research, promote findings, and train application. As described in the FAA Order 8040-4: "This order establishes the safety risk management policy and prescribes procedures for implementing safety risk management as a decision-making tool within the Federal Aviation Administration (FAA)." (FAA, 1996, p. 1)

This Administrative Order requires the Office of System Safety to: 1) incorporate a risk management process for all high-consequence decisions, and 2) provide a handbook/manual of System Risk Management and to recommend tools of System Safety to all US-based airlines (FAA, 1996). To accomplish the appointed tasks, an annual System Safety conference and workshop available for airline managers has become routine since 1999. The research efforts from the FAA, project contractors, or conference participants were exchanged and ideas were discussed during each workshop. Despite the handbook of System Safety containing essential System Safety theories, the current System Safety studies from the industry are limited to engineering design such as navigation system; weather and turbulence forecast; global positioning systems; runway incursion; consumer safety guidelines; and airport operational procedures.

Lu, Wetmore, and Przetak (2006) conducted a Content Analysis study of System Safety. They discovered that the FAA's advocate was mostly concerned with the conceptual procedure of risk management, rather than an in-depth demonstration or usage of safety analysis techniques (see Appendix A). As a result, most airlines (flag or non-flag) and FBOs did not implement non-mandatory training, such as Maintenance Resource Management (MRM) or System Safety management to its operation unless a voluntary engagement (i.e., Air Transportation Oversight System [ATOS] or Advanced Quality Program [AQP]) had been initiated (Lu, Przetak, & Wetmore, 2005). In addition, the use of System Safety concepts has primarily been tied to risk management using a basic descriptive trend study on a voluntary basis. Examples of such voluntary programs include ATOS, AQP, FAA Safety Reporting System and Database (SRSD), NASA Aviation Safety Reporting System (ASRS), Flight Operational Quality Assurance (FOQA), Air Carrier Operations System Model (ACOSM), and American and Delta Airlines' Aviation Safety Action Program (ASAP). All the aforementioned safety programs are trend study, checklist type, hazard-identification oriented but they are segregated instead of integrated into one system. This situation has been addressed by the Government Accounting Office (GAO) with a special report, *System Safety Approach Needs Further Investigation into FAA's Oversight of Airlines*. The report assesses the FAA's strengths and weaknesses of the inspection approach to the non-legacy airlines and questions the effectiveness of the System Safety program. The results indicated: 1) the strength of System Safety management is the prioritization of risk containing hazard probability (P) and severity (S) (USGAO, 2005a) no different to the ongoing advisory program for flag airlines, and 2) the voluntary risk management embracing System Safety concept can dramatically assist the FAA in reducing workload, because inspection data has been compiled into prioritized data banks by airlines up front. Nevertheless, airline employees did not grasp the full benefit of implementing System Safety. In some cases, the information was not explained clearly; however, in many instances aviation workers cannot locate the needed information. Incomplete information resulted in the erroneous allocation of valuable and scarce resources to non-risk or low-risk tasks from the airlines as well as the FAA. The report also mentioned the FAA safety inspectors' knowledge and technical capability in Systems Safety should also be further improved (USGAO, 2005b). These details indicate an opportunity for improvement and suggest a more comprehensive, user-friendly, and dynamic system safety model.

FAA AC 120-92

In 2006, the FAA published its Advisory Circular (AC 120-92) *Introduction to Safety Management Systems for Air Operators* to meet two goals: 1) introducing the concept of safety management system to air transportation service providers, and 2) providing air carriers a guideline of safety management system (SMS). The merits of this AC focus on: 1) safety management (risk management, safety assurance using quality management techniques, and systemic approach of safety management), and 2) safety culture (the human-centered psychological, behavioral, and organizational elements) (FAA, 2006). Using safety risk management and safety assurance to manage safety in this publication is plausible. However, the model embracing risk matrix given by the FAA lacks insightful details. First, user-friendly format of report forms is not covered, which creates complexity of the model. Secondly, system safety tools like Fault Tree Analysis (FTA), Operations, and Support Hazard Analysis (O&SHA) should be recruited to this safety guideline so aviation industry could have a bigger picture for implementing system safety management.

The MIL-STD-882D and Policymaking Challenge

The original intent of the *Standard Practice for System Safety* (MIL-STD-882A published in 1969) was to help aircraft and aerospace engineers better design products without utilizing expensive fly-fix-fly doctrine embraced by the U.S. military (especially the early project teams of X-planes) before and directly after World War II. After 1969, the U.S. Air Force and NASA both realized that the MIL-STD-882 was extremely helpful in reducing system active and latent failures (DOD, 2000).

In fact, an accident is rarely caused by a singular hazardous factor or an isolated risk (Petersen, 1988; Wells, 2004; Wood, 2003). This theme can be examined by traditional safety models such as 5-M factors, Swiss Cheese model, Domino effect, SHELL model, Chain-of-events and related safety analysis mechanisms. Therefore, the concept of multi-factor *causes* → *event* ($X_s \rightarrow Y$), where multiple causes contribute to the accident, is not a contentious issue. The cause, X, could be identified as a violation, distraction, complacency, carelessness, recklessness, fatigue, poor situational awareness, etc. Although each is considered one element, there are numerous precursors to the so-called single element providing multiple factors causing Y. For instance, the cause of the recent accident of Comair Flight 5191 (initiating takeoff on the wrong runway) could be human error concerning miscommunication, situational awareness, crew resource management, flight training, ATC's breach of procedure, or other causal factors. Nevertheless, the most reasonable question is: Why did the human (pilots or air traffic controller) error occur in the first place? Was it due to training, personal problems, health, sociopsychological status, or simply operational carelessness?

Problems with management can challenge organizational safety. In 1997, James Reason published a book, *Managing the Risk of Organizational Accidents,* showing that people make mistakes no matter what their intentions are. Reason categorized human behavior into three subgroups: 1) skill-based behavior (SB), 2) rule-based behavior (RB), and 3) knowledge-based behavior (KB). To identify potential errors hidden in the dark corners of a given management system, Reason (1997) suggested the examination of salient problems by three dimensions: 1) Personal, a worker is an agent of a system, 2) Engineering, statistical prediction, and 3) Organizational, different management segments of an organization are responsible for safety. The three models intertwine. For example, the organizational model contains both the personal and engineering models. Thus understanding an organizational accident is critical to safety performance. For instance, an accident could occur when a breakdown prevails within a hierarchical management system; and in particular, when a safety management structure is ill-formed (Reason, 1997).

In his book, *Safety and health: Management planning,* and his paper, *Three Ps in Safety: Policies, Procedures, & Performance*, Ted Ferry (1990; 2006) emphasized the essentiality of setting up a policy so procedure is created and performance is measured. Moreover, the 3Ps concept is not a linear process per se, but a recursive and cyclic activity linking policies, procedures, and performance into one frame (see Figure 1).

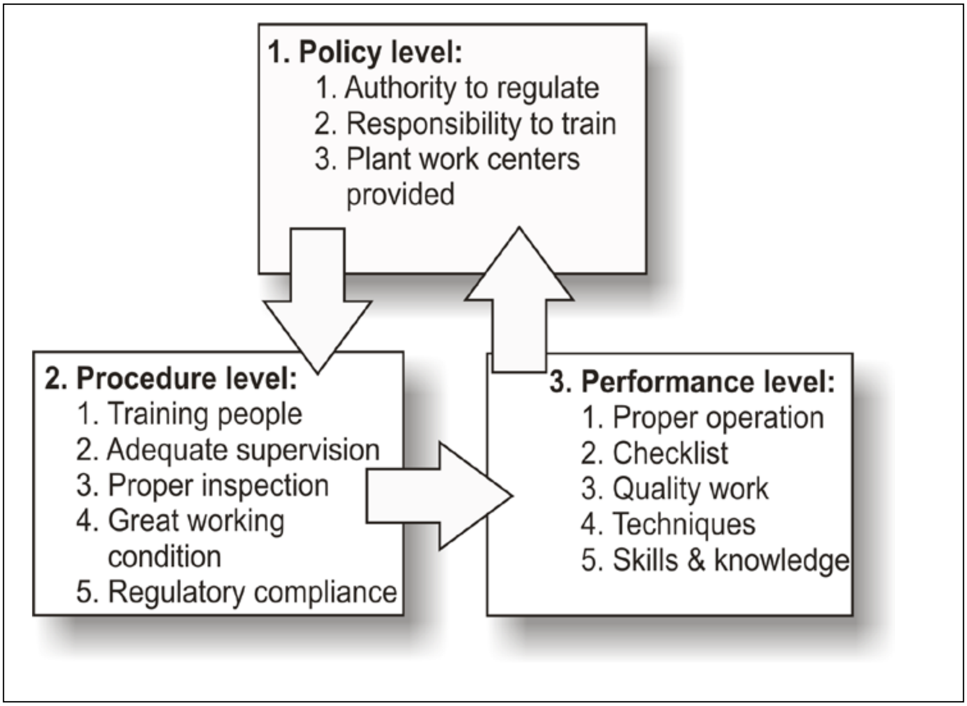


Figure 1. The correlation among 3Ps

If any segment of a management system is lacking, the entire safety loop will collapse and the program will fail. Apparently, setting policy is needed and should be first done by the government. This concept has been echoed by Wells (2001) stating that “Policy is usually referenced in the early part of all plans” (Wells, 2001, p. 234). However, in today’s aviation industry, establishing a new policy is not only unwelcome by the industry, but could also be extremely time-consuming without immediate threats (Lu, 2005). An alternative is to develop a foundation for working procedures so all employees can follow proper guidelines and achieve safety performance (Reason, 1997; Lu, 2005).

Definitions of System Safety Techniques

To proactively identify potential hazards leading to human errors and accidents, MIL-STD-882D suggests several techniques that System Safety experts can apply. The System Safety techniques described below are the most commonly used.

> *Job Safety Analysis (JSA).*
> A generalized examination of the tasks associated with the performance of a given job and an evaluation of the hazards associated with those tasks and the controls used to prevent or reduce exposure to those hazards. Usually performed by the responsible supervisor for that job and used primarily to train and orient new employees (Vincoli, 2006, p. 206).

> *Operating & Support Hazard Analysis (O&SHA).*
> Performed to identify and evaluate operational-type hazards. It is based upon detailed design information and is an evaluation of operational tasks and procedures. It considers human system integration factors such as human error, human task overload, cognitive misconception, [and] the effect on humans of hardware failure (Ericson, 2005, p. 476).

> *Fault Tree Analysis (FTA).*
> Systems analysis technique used to determine the root causes and probability of occurrence of a specified undesired event. A fault tree analysis is a model that logically and graphically represents the various combinations of possible events, faulty and normal, occurring in a system that lead to a previously identified hazard or undesired event (Ericson, 2005, p. 472).

> *Failure Mode and Effective Criticality Analysis (FMECA).*
> Tool for evaluating the effect(s) of potential failure modes of subsystems, assemblies, components, or functions. It is primarily a reliability tool to identify failure modes that would adversely affect overall system reliability. FMECA has the capability to include failure rates for each failure mode in order to achieve a quantitative probabilistic analysis (Ericson, 2005, p. 471).

> *Management Oversight and Risk Tree (MORT)*. "MORT is an analytical technique for identifying safety-related oversights, errors, and/or omissions that lead to the occurrence of a mishap" (Ericson, 2005, p. 423). Regardless of the informed system risks, the focus of this model is on management's omissions or less than adequate (LTA) performance.

Recent Studies of System Safety Management

In addition to aerospace engineering, applications of System Safety techniques and concepts have been useful in the fields of medicine and aviation safety. For instance, in the medical engineering industry, Helmreich (2000) advocated the use of the System Safety's error management concept in medical practice. In 1999, other related research was done by Croheecke and his research associates using FMEA (Cromheecke, Mak, & de Mol, 1999). Hyman (2002) utilized the leading tool of System Safety, the FTA, in evaluating potential hazards associated with newly innovated medical devices before moving toward the production-manufacturing phase in the device's life cycle.

In aviation safety, the U.S. Air Force launched risk management and causal study to improve pilot training procedures. Diehl's (1991) cross-referenced analysis of 208 military accidents discovered the breakdown of cockpit communication and team performance known as crew coordination. This communication breakdown directly led to military aircraft mishaps. Diehl's study applied ergonomic human-face interface and suggested a modification of the cockpit layout of the Cessna Citation used by U.S. Air Force officers. His study linked accident investigation, hazard identification, and basic descriptive analysis to human factor and CRM training. Lu, Przetak, and Wetmore (2005) did a similar causal study using statistical analysis to discover the causes of non-flight accidents for FAR Part 121 U.S.-based carriers.

Another study by Thom and Clariett (2004) published in Collegiate Aviation Review (CAR) focused on an essential section of System Safety, the applicability of job safety analysis (JSA) and task analysis. In their study, JSA was closely interpreted and the layout of human-machine interface was emphasized. Using the Risk Homeostasis Theory (RHT) of human dynamic behavior on risk taking, their study helped identify potential hazards surrounding the hangar, factory, or student workshop both internally and externally (Thom & Clariett, 2004). This study was of great interest to the maintenance safety community.

Bastos (2005) presented a risk management model based on the feedback from 14CFR Part 135 pilots as well as NTSB's accident data. Bastos discovered that accidents are usually caused by multiple safety factors concurring with Reason's theory of organizational accidents. Since potential risk exists, reducing risk probability (Rp) and risk severity (Rs) upstream is essential. Bastos also proposed a risk management model similar to MIL-STD-882D Risk Matrix (See Appendix B). However, his model is specifically tailored for air taxi operation.

Lu, Wetmore, and Przetak (2006) demonstrated FTA application in promoting safety performance. Their analysis indicated that FTA, based on a risk tree analysis and statistical forecast, can help to proactively prevent undesired events or stop events from happening. In their study, FTA helps organizations such as the government or airlines effectively and promptly identify accident postulates that lead to implementation of strategic safety prevention programs from the bottom-up (upstream) (see Appendix C). Based on the FTA hierarchical block-diagram in their study associated with Boolean gates, any of the root factors on the bottom level can form a cut-set or a failure-chain, which can result in an accident or system failure (top event). Hence, compressing or eliminating the failure probability of root factors from the lowest level of the risk tree is a training priority. A statistical simulation based on the required risk calculation concerning hazardous probability and severity is shown in Appendix C. With the computerized-system design, a real-time and dynamic System Safety model is possible.

Although the FAA realized the value of collecting data and monitoring trends, the FTA model provides a dynamic system to identify hazards and to assign risk values using Bayesian analysis. Bayesian inference provides a quantitative framework for the iterative process of integrating information into useful models. It is the essence of predicting and therefore preventing accidents. The model is the keystone to system safety and may have been unnoticed by the FAA. Since it was not obligatory, the aviation industry did not use it.

Lupolis (2006) from the Brazilian Air Force accomplished the recent application of Organizational Accident theory in July 2006. His research "Discovering the Commanders' Perception Regarding Organizational Accidents" hypothesized that deficient decision-making processes and poor organizational management from top-ranking officers could result in aircraft mishaps. He revealed, via self-administrated surveys, that Brazilian Air Force squadron commanders have a limited knowledge about advanced safety theories like the Organizational Accident theory, but they are all committed to operational safety. As a result, a more advanced safety education for top management is needed. Furthermore, the top-ranking officers are aware of their lack of knowledge about a reliable decision-making process. They still use empirical means to make safety decisions. Lupolis recommended further research on the rationale these officers use to make safety decisions. Is the correction one-way, downstream, or is there an iterative process, moving both upstream and downstream?

Safety related courses and theories are provided at U.S. based universities offering aviation education, so students can practice not only as professional aviators, but also as safe ones. At the University of Central Missouri, the traditional safety theories and models taught by the Department of Aviation include: a) SHELL model, b) Domino Effect, c) Swiss Cheese model. d) Chain-of-events, e) Human Behavior, f) Ergonomics, g) Risk Management and Error Reduction, h) Hazardous Materials and Environment, i) Aeronautical Decision Making, j) 5-M model, and k) Hazardous Attitudes. Most academic aviation programs include safety programs advocated by the FAA. These include programs such as Flight Operational Quality Assurance (FOQA), Aviation Safety Reporting System (ASRS), Aviation Safety Action Program (ASAP), Air Transportation Oversight System (ATOS), Advanced Quality Program (AQP), and the International Aviation Safety Assessment (IASA). Despite the academic education on campus in relation to a student's career focus, the Department of Aviation also works closely with the FAA Kansas City Flight Standard District Office (FSDO) to enhance student's knowledge on safety thru periodical training, seminars, and a series of guest lectures. Meanwhile, the faculty work as safety counselors thru programs such as FAASTeam. These initiatives help the FAA's program of *Safety through Education* become more successful in the Central region. Yet the ongoing safety programs using System Safety are individualized, separated, and not unified for easy usage.

Research Focus

The FAA is now intensively advocating the adoption of System Safety by the aviation industry. Although related studies are limited and sporadic, aviation safety managers are encouraged to utilize System Safety concepts and techniques in auditing aviators' daily activities in a hope that an internal safety culture can be formulated. To this end, a safety model comprehensively outlining the often overlooked analytical procedures and system safety techniques will be beneficial to the aviation community.

Research Methodology

The research methodology applies qualitative case study to demonstrate the functionality of the proposed safety management model. A case is the object of study, namely the unit of analysis. The unit of analysis may be a narrative report, a technically distinctive situation, an event, a location, a unique organization, or even a suggested resolution. With this in mind, conducting a case study is a systemic analysis followed by a proposal for an alternative solution of the situation (Berg, 2001; Yin, 1994). The case selected in this paper is “Federal Express Flight 1478” from the NTSB. The investigation was thorough and the results were voluminous which will, to a certain extent, ensure internal validity (the accuracy of measurement of phenomenon or object by research tools and associated procedures). The external validity (the magnitude of result generalization reflecting genuine needs of the study), of the proposed safety model must meet the following criterion: 1) administratively practical, 2) the basis of measurement should be quantifiable for qualitative analysis, 3) a valid measurement must present what it is supposed to represent, 4) system safety tools utilized should be understandable, user-friendly and sensitive to situational change, 5) safety data should be in a real-time reflection/alert fashion, and 6) the safety result should be distributable and disseminated (U.S. Army, 1972; Wood, 2003).

Proposed Aviation System Safety Management Model (ASSMM)

The proposed Aviation System Safety Management Model (ASSMM) contains nine (9) major steps: 1). Data Collection, 2) Hazard Identification, 3) Data Analysis, 4) Risk Matrix Calculation & Response, 5) System Safety Tools implementation & Regulatory Compliance, 6) Reports & Feedback, 7) Result Monitoring, 8) Information Distribution, and 9) Problem-solving meeting (See Figure 2).

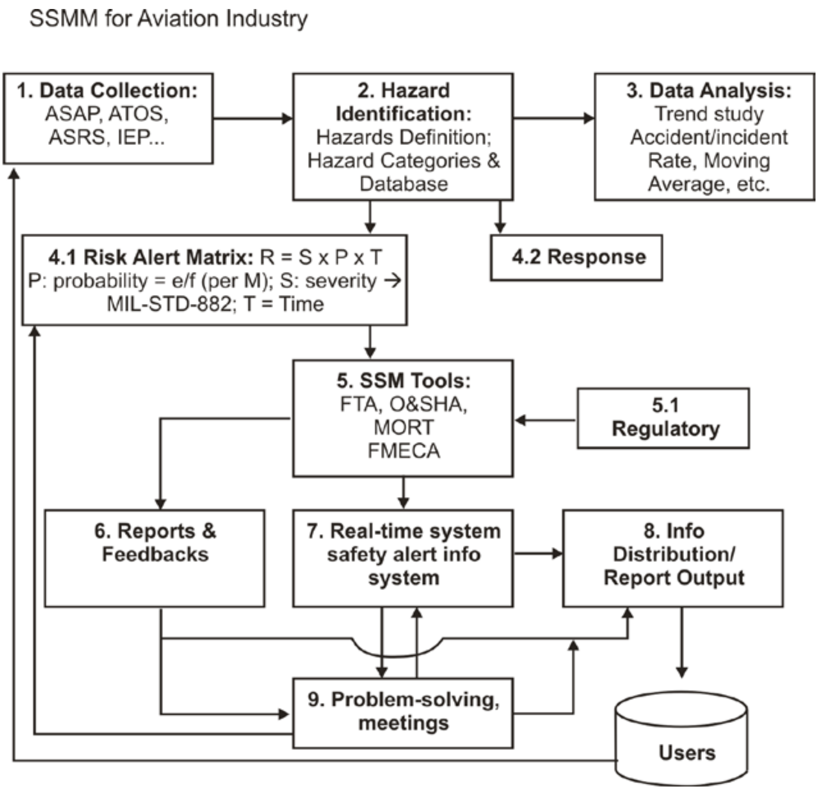


Figure 2. SSMM for Aviation Industry

Data collection.

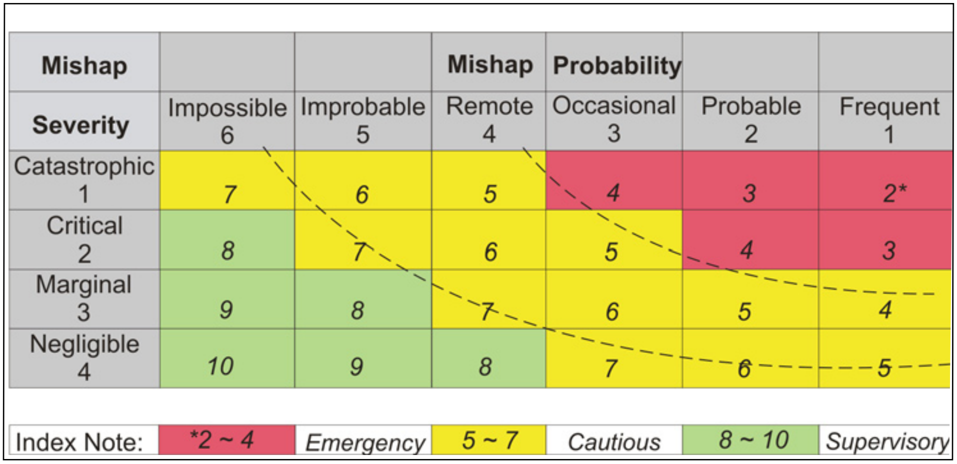
Hazardous data can be retrieved from the current ongoing reporting programs such as ATOS, ASRS, ASAP, AQP, and FOQA automatically. The data collection process can be: 1) reported by employees, 2) downloaded from a Flight Data Recorder (FDR), or 3) a documentary review. A suggestion of this data collection phase is that the data reporting mechanism should be open to all workers allowing safety project managers to collect sufficient information from field specialists. This collection must meet several requirements in order to encourage contributions: 1) penalty-free, 2) anonymous, 3) confidential, 4) easy-to-report, 5) maintaining an open-door policy, and 6) promising solutions.

Hazard identification. The purpose of hazard identification is twofold: hazard definition and data categorization. The criticality of hazard identification focuses on the review of hazard reports from line experts to see if it is a reportable hazard and requires further analysis. In addition, collected data should be categorized and prepared for a prompt analysis and immediate hazard study.

Data analysis. This is the first analytical output of review focused on identifying and reporting hazard prioritization associated with a quick solution or immediate safety alert. Data analysis should contain but not be limited to some basic hazardous information such as trend study, hazard ranking, and preliminary accident and/or incident rates during a specific time.

Risk matrix calculation and response. During this phase of system safety management model (ASSMM), the formation of a Risk Index Matrix can be flexible as long as a precise risk index can be generated. An example shown in Table 1, the Risk Index Matrix embraced addition instead of multiplication providing an easier way of risk calculation and interpretation ranging between 2 and 10, the higher the better.

Table 1
ASSMM Risk Index (ASSMMRI)



Mishap			**Mishap**	**Probability**		
Severity	Impossible 6	Improbable 5	Remote 4	Occasional 3	Probable 2	Frequent 1
Catastrophic 1	7	6	5	4	3	2*
Critical 2	8	7	6	5	4	3
Marginal 3	9	8	7	6	5	4
Negligible 4	10	9	8	7	6	5
Index Note:	*2 ~ 4	Emergency	5 ~ 7	Cautious	8 ~ 10	Supervisory

Meanwhile, a color-coded numerical index matrix indicates the risk level of a situation reported by employees. In this proposed model, the risk index (2~4) is qualitatively defined as an Emergency case that needs an immediate response or solution. The risk index (5 ~7) indicates a Cautious situation needing a fast review and resolution, which more information and analysis may be needed to determine the level of risk eroding the entire operational system. Lastly, the risk index (8 ~10) represents a Supervisory case and the reported hazard needs a continuous measurement in the future. In the above matrix for the aviation industry, although the mishap probability is extremely low (Impossible 6), any fatality (Catastrophe 1) is unacceptable thus it is categorized as Cautious instead of Supervisory. Meanwhile, Frequent (1) mishap probability with Negligible (4) hazard severity is also unacceptable because the hazard could be immediately mitigated with a very low cost (i.e., tool misplacement) otherwise hazard accumulation (i.e., complacency) may lead to a larger scale of damage.

System safety tools implementation and regulatory compliance. This phase processes the information/reports and receives the hazard probability from previous processing stage. The accident case of “Federal Express Flight 1478” was selected using FTA (See Appendix E) and O&SHA for a conceptual demonstration (See Appendix F). The genuine value of this phase is to apply System Safety tools to conduct a detailed hazard-accident-incident analysis and provide counter-measures for new employee orientation, routine safety education, recurrent training, and an accident-prevention course based on regulatory requirements and identified safety gaps within the operational system.

Reports and feedback. The purpose of accident investigation is to identify the problems, provide safety measures, and prevent similar problems from happening again. With this in mind, the analytical reports will be sent to a safety committee for review if the calculation of Risk Index indicates a need. Also, the result and resolution needs to be distributed to the submitters, if known. Otherwise, it should be posted on a service bulletin board or to a monitoring system for public review. A hazard tracking system is equally important for two counts: 1) it will help the safety manager identify the status of a hazard report, and 2) it will show hazard submitters the importance of their report and further motivate their participation.

Real-time system safety alert. Qualitative risk alert index of this proposed ASSMM provides a visionary image to safety managers or system users who need up to date information for prompt understanding. The authors suggest a color-coded (red, yellow, green) information distribution design for informative hazard identification.

Information distribution. This process should inform all employees. It is accomplished by utilizing several formats such as email, voicemail, internal circulation channel, flight crew briefings, ground crew discussions, maintenance safety notices, and recurrent and routine training.

Problem-solving meeting. Members of safety committee receive routine, at least daily, hazard analysis and provide comments and recommendation to upper management for further decision-making reviews (action or non-action). Line managers or employees should be invited to meetings and suggest training or resolution if deemed necessary.

Conclusion and the Challenge of System Safety

It took almost 30 years for the FAA to recognize the value of System Safety concept (from hazard identification, risk analysis, safety countermeasure, performance assessment, and product documentation); the recognition, unfortunately, is rarely expended to its key analytical techniques such as Job Safety Analysis (JSA), Operating & Support Hazard Analysis (O&SHA), Fault Tree Analysis (FTA), Failure Mode and Effective Criticality Analysis (FMECA), and Management Oversight and Risk Tree (MORT). As a result, the airline who implements specific System Safety techniques to their operation is rare.

Over the last seven years, airline industry and the FAA have both argued that a sufficient database is critical to a mature risk prediction, mishap mitigation, and policy making. With this in mind, the reluctance of further using System Safety techniques by the air carriers is not only arguable but also dangerous to its loyal passengers. Without a doubt, System Safety techniques specifically follow a systemic and scientific way to trouble-shoot a given system by identifying potentials hazards endangering the whole operating system within the error-latent environment.

This paper provides a comprehensive safety model, namely Aviation System Safety Management Model (ASSMM), combining Error Management, System Safety techniques and MIL-STD-882D to form a streamline risk analysis and accident prevention program for easier usage and safety reporting. A case study was applied to demonstrate the fundamental application of this newly proposed ASSMM. Because this is a conceptual safety model, readers can freely, with flexibly, revise this model using different system safety techniques based on the characteristics of a new hazardous environment. However, the foundation of any new safety analysis and prevention models could be rooted in this proposed ASSMM.

Future Study

Using the computer to solve problems is the future. Advocating a follow-up study implementing FTA, FMECA or MORT by designing a “computerized” statistical analysis and risk prioritization shall improve discipline of predicting hazards and generate an automatic auditing or alert model for a system’s real-time operational safety. Another follow-up study should focus on performance and usefulness of this proposed ASSMM tested by airlines or FBOs.

References

Bastos, L. C. M. (2005).*Risk management model for on-demand Part 135 (Air Taxi) operators.* Master Thesis. Warrensburg, MO: University of Central Missouri.

Berg, B. (2001). *Qualitative research methods for the social sciences* (4th ed). Boston: Allyn and Bacon.

Cromheecke, M. E., Mak, R., & de Mol, B. A. J. M. (1999). Failure mode & effect analysis and fault tree analysis in the use of the CoaguChek® prothrombin time system. *International Journal of Risk in Medicine*, *12*(3-4), 173-179.

Department of Defense [DOD] (2000). *MIL-STD-882B: System Safety Program Requirements.* Author.

Department of Defense [DOD] (2000). *MIL-STD-882D: Standard practice for system safety*. Author.

Diehl, A. E. (1991). Human performance and systems safety considerations in aviation mishaps. *International Journal of Aviation Psychology, 1*(2), 97-106.

Ericson, C. A. III (2005). *Hazard analysis techniques for System Safety*. Hoboken, NJ: Wiley & Sons.

FAA. (1996). *FAA Order 8040-4 Safety Risk Management*. Author.

FAA. (2000). *System safety handbook*. Author.

FAA. (2006). *AC 120-92 Introduction to safety management systems for air operators*. Author.

Ferry, T. S. (1990*). Safety and health: Management planning*. New York: Van Nostrand Reinhold.

Ferry, T. S. (2006). Three Ps in safety: Policy, procures, and performance*. Professional Safety, 51*(6), 48-52.

Helmreich, R. L. (2000). On error management: Lessons from aviation. *British Medical Journal*, *320*(7237), 781-785.

Hyman, W. A. (2002). Generic fault tree for medical device error, *Journal of Clinic Engineering, 27*(1), 134-140.

Lu, C-t. (2005). Discovering the Regulatory Considerations of the Federal Aviation Administration: Interviewing the Aviation Rulemaking Advisory Committee, *Journal of Air Transportation, 10*(3), 32-48.

Lu, C-t,, Przetak, R., & Wetmore, M. (2005 June). Discovering the non-flight hazards and suggesting a safety training model. *International Journal of Applied Aviation Science, 5*(1), 135-152.

Lu, C-t., Wetmore, M., & Przetak, R, (2006). Another approach to enhance airline safety: Using System Safety techniques. *Journal of Air Transportation, 11*(2), 113-139.

Lupolis, L. C. (2006). *Discovering the Brazilian Air Force squadron commanders' perceptions regarding organizational accidents*. Master Thesis. Warrensburg, MO: University of Central Missouri.

National Commission on Terrorist Attacks. (2004, July). *The 9/11 commission report: Final report of the National Commission on Terrorist Attacks Upon the United States* (Authorized Edition). New York: W.W. Norton.

NTSB. (2002).*Collision with trees on final approach: FedEx Flight 1478*. Author.

Petersen, D. (1988). Safety management: A human approach (2nd. ed.). Goshen, NY: Aloray.

Reason, J. (1997). *Managing the risks of organizational accidents*. Burlington, VT: Ashgate.

Thom, J. M., & Clariett, J. R. (2004). A structured methodology for adjusting perceived risk, *Collegiate Aviation review, 22*(1), 97-121.

US Army (1972). *Safety program*. Author.

USGAO. (2005a). *System Safety approach needs further integration into the FAA's oversight of airlines*. Author.

USGAO. (2005b). *FAA's safety oversight system is effective but could benefit from better evaluation of its programs' performance*. Author.

Vincoli, J. W. (2006). *Basic guide to System Safety* (2nd ed.). Hoboken, NJ: Wiley & Sons.

Wells, A. T. (2004). *Commercial aviation safety* (4th ed.). New York: McGraw Hill.

Wells, A.T., 2001: Commercial aviation safety. 3rd Edition, McGraw-Hill, New York, NY.

Wells, A.T. & Rodrigues, C.C. (2004). *Commercial Aviation Safety (Fourth Edition).* New York: McGraw-Hill.

Wood, R. H. (2003). *Aviation safety programs: A management handbook* (3^{rd}. ed.). Englewood, CO: Jeppesen.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Beverly Hills, CA: Sage Publishing

Appendix A

SYSTEM SAFETY WORKSHOPS AND CONFERENCES – CONTENT ANALYSIS

	2001	2002	2003	2004
System Safety Management	X	X	X	X
Aviation System Safety Program (AvSP)	X	X	X	X
FAA-Airlines Collaboration	X	X	X	X
Data Collection & Risk Analysis	X	X	X	X
System Risk Management (SRM) & Safety Culture		X	X	X
Flight crews-centered	X	X		X
Non-flight crews-centered	X	X	X	X
All aviation workers	X			
Air Carrier Operations Systems Model (ACOSM)	X			
Aviation Safety Action Program (ASAP)	X	X		X
Flight Operational Quality Assurance (FOQA)	X	X		X
Advanced Quality Program (AQP)	X			
Aviation Safety Reporting System (ASRS)	X			X
Continuous Analysis and Surveillance System (CASS)	X			
Maintenance Resource Management (MRM) training	X	X		
Human Factor CRM training	X	X	X	X
Cased-based training/Naturalistic Decision-making	X	X	X	X
Regulations	X	X	X	
Cost-benefit and Safety Investment	X	X	X	X
Failure Mode and Effective Analysis (FMEA) Concept		X		
Failure Mode and Effective Analysis (FMEA) Application				
Fault Tree Analysis (FTA) Concept		X		
Fault Tree Analysis (FTA) Application				
Risk Control Management (RCA)				X
Hybrid Causal Modeling			X	X

Note: The origin of this Content Analysis Table was statistically extracted from the research projects and papers presented at the FAA System Safety workshops and conferences between 2000 and 2004. As shown in the above table, most researches either focused on the advocate of using System Safety concepts or risk analysis covering trend study. Researchers did not apply tools (i.e. FTA or FMEA) to their studies for demonstration. Especially, there were only two papers that explained FMEA and FTA techniques over the past four years. Yet no further application was found.

Appendix B

RISK MATRIX, SEVERITY & PROBABILITY

Risk Matrix

Frequency	Catastrophic (I)	Critical (II)	Marginal (III)	Negligible (IV)
Frequent (A)	1A	2A	3A	4A
Probable (B)	1B	2B	*3B*	*4B*
Occasional (C)	1C	*2C*	*3C*	4C
Remote (D)	*1D*	2D	3D	4D
Impossible (E)	1E	2E	3E	4E

*A "Risk" falling into this category [1A, 2A, 3A, 4A, 1B, 2B, 1C] is "Unacceptable"
A "Risk" falling into this category [1D, 2C, 3B, 3C, 4B] is "Undesirable"
A "Risk" falling into this category [1E, 2D, 2E, 3D, 4C] is "Acceptable with Review"
A "Risk" falling into this category [3E, 4D, 4E] is "Acceptable without Review"
The determination of "Unacceptable," "Undesirable," "Acceptable with Review," or "Acceptable without Review" is based on a System Safety analyst's subjective decision-making based on the onsite situation from case to case.

Risk Severity (S) and Probability (P) are defined as:

Risk Severity (S)

Description	Category	Mishap Definition
Catastrophic	I	Death or system loss/failure
Critical	II	Severity injury, occupational illness, or system damage
Marginal	III	Minor injury, occupational illness, or system damage
Negligible	IV	Other

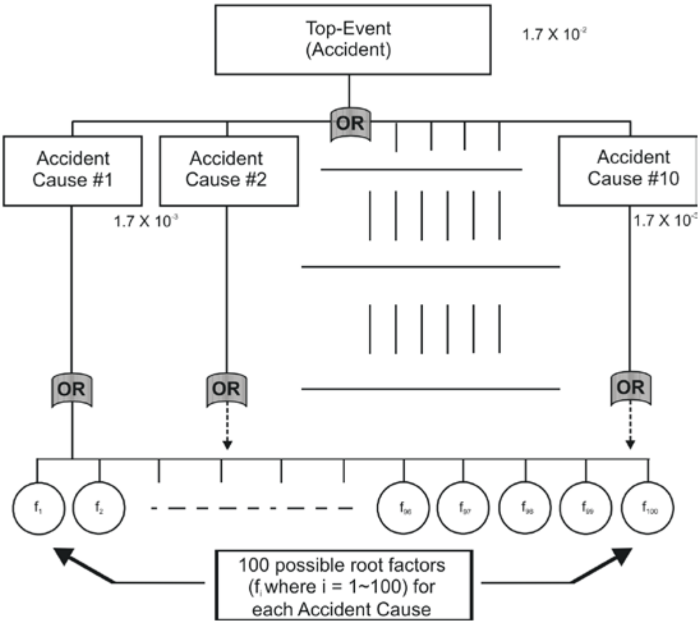
Risk Probability (P)

Description	Level	Mishap Definition
Frequent	A	Likely to occur frequently
Probable	B	Will occur several times during the life of an item
Occasional	C	Likely to occur sometimes in the life of an item
Remote	D	Unlikely, but may possibly occur in the life of an item
Impossible	E	So unlikely, assumed that hazard will not occur at all

Source: DOD MIL-STD-882B System Safety Program Requirements (1984)

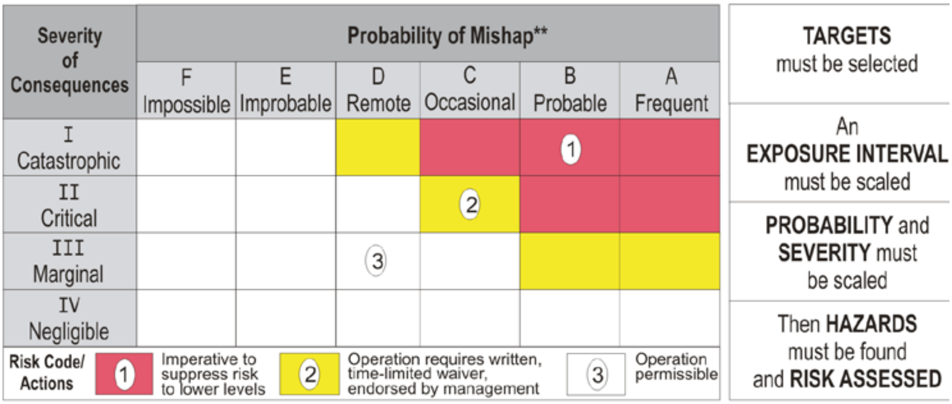
Appendix C

SIMULATING THE PROBABILITY OF THE TOP-LEVEL EVENT



Appendix D

Risk Index Matrix



Severity of Consequences	Probability of Mishap**					
	F Impossible	E Improbable	D Remote	C Occasional	B Probable	A Frequent
I Catastrophic			🟨	🟥	🟥 ①	🟥
II Critical				🟨 ②	🟥	🟥
III Marginal			③		🟨	🟨
IV Negligible						
Risk Code/ Actions	🟥 ①	Imperative to suppress risk to lower levels	🟨 ②	Operation requires written, time-limited waiver, endorsed by management	③	Operation permissible

TARGETS
must be selected

An
EXPOSURE INTERVAL
must be scaled

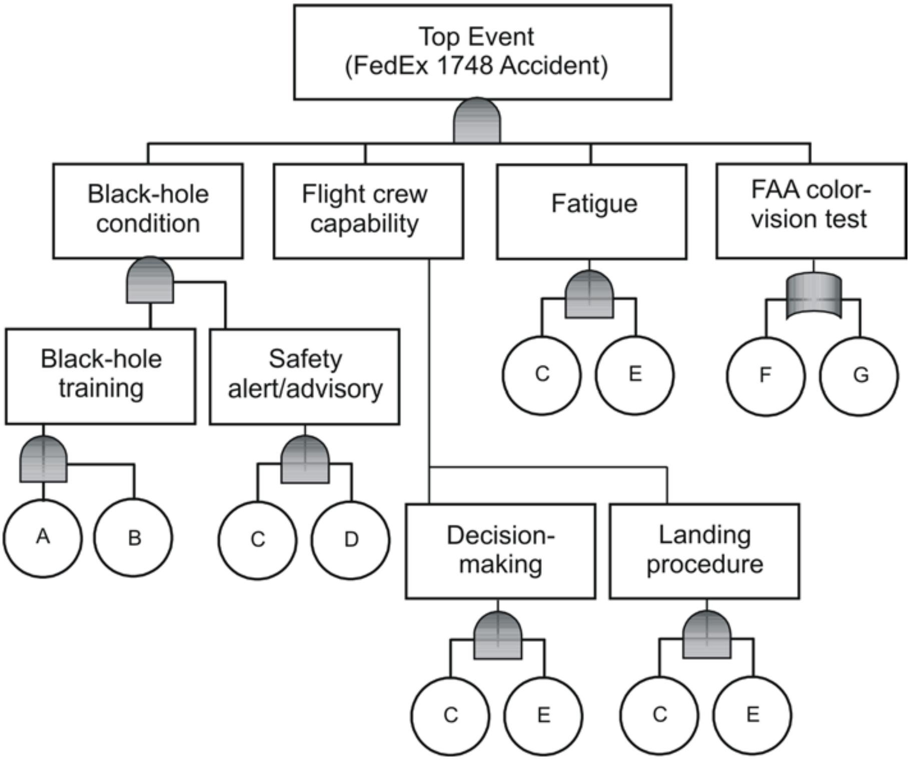
PROBABILITY and
SEVERITY must
be scaled

Then **HAZARDS**
must be found
and **RISK ASSESSED**

*Adapted from MIL-STD-882 **Life Cycle = 25 yrs

Source: MIL-STD-882D (DOD, 2000) and FAA System Safety handbook (FAA, 2000, p. 3-8).
Note: The scale of "Probability of Mishap" ranged from "A" (Frequent), "B" (Probable), "C" (Occasional), "D" (Remote), "E" (Improbable), and "F" (Impossible). Based on the FAA's System Safety Manual, the associated mishap probabilities (P) are recommended as: A ($A>10^{-5}$), B ($10^{-5}>B \geq 10^{-7}$), C ($10^{-7}>C \geq 10^{-9}$), D ($10^{-9}>D \geq 10^{-11}$), E ($10^{-11}>E \geq 10^{-13}$), F ($F<10^{-13}$); while "Mishap Severity" is defined as shown in Appendix B of this paper. Researchers should tailor and define his own operational mishap probability and severity based on the nature of its operation (i.e., Part 121, Part 135, General Aviation, FBO, etc.).

Appendix E
FedEx Flight 1478 – Conceptual FTA Analysis



<u>Note:</u>

> The National Transportation Safety Board determines that the probable cause of the accident was the captain's and first officer's failure to establish and maintain a proper glide-path during the night visual approach to landing. Contributing to the accident was a combination of the captain's and first officer's fatigue, the captain's and first officer's failure to adhere to company flight procedures, the captain's and flight engineer's failure to monitor the approach, and the first officer's color vision deficiency (NTSB, 2002, p. 68).

<u>Definition of root causes:</u> "A": Initial Training *LTA*, "B": Recurrent Training *LTA*, "C": Airline Advisory *LTA*, "D": Airport NOTAM *LTA*, "E": Flight Capability *LTA*, "F": FARs *LTA*, "G": Vision checkup *LTA*. "*LTA*": Less Than Adequate.

Appendix F
FedEx Flight 1478 – Conceptual O&SHA Analysis

Item	Procedural Tasks	Hazard condition	Cause	Effect	Risk level, Criticality index	Assessment	Recommendation
1	Receive weather information	Inaccurate information	Communication deficiency	Poor judgment	IIIB, 5 (cautious)	Operator error	Assurance of information accuracy, CRM
2	Landing runway	Configure and condition	Incorrect information	Poor judgment	IIIB, 5 (cautious)	Operator error	Assurance of information accuracy, CRM
3	Decent	Info accuracy	ATC error	Abnormal decent	IIC, 4 (emergency)	Operator error	Situational Awareness, CRM
4	CFIT/GPWS	Pilot awareness	Crew error	Judgment	ID, 4 (emergency)	Operator error	FAA approved SOPs
5	Approach briefing	Pilot awareness	Crew error	Judgment	IIB, 4 (emergency)	Operator error	FAA approved SOPs, CRM
6	Aircraft system check	System failures	Pilot capability & awareness	Crew error	Judgment & skills	IIB, 4 (emergency)	FAA approved SOPs, Ground school enhancement, Aircraft type rating recurrence
7	Airport condition checkup	Unprepared ground condition	Communication deficiency	Abnormal landing	IIB, 4 (emergency)	Operator error	FAA approved SOPs, Airport advisory info, NOTAM
8	Pre-landing check	Omission	Lack of training, hazardous attitude	Poor landing result	IIIB, 5 (cautious)	Operator error	SOPs; Situational Awareness, CRM
9	Visual approach	Non-clear path	ATC Communication deficiency	Dangerous situation	IIC, 4 (emergency)	Operator/ATC error	Situational Awareness; ATC/Crew training
10	PAPI light	Indicator clarity	Crew/system error	Inaccurate flight operation	IIC, 4 (emergency)	Operator/ATC error	Situational Awareness; Crew training; System inspection

Synopsis of the accident: On July 26, 2002, FedEx flight 1478 struck trees on short final approach and crashed short of runway 9 at the Tallahassee Regional Airport (TLH), Florida. The three flight crews were seriously injured, and the airplane was destroyed by post-crash fire. The O&SHA concerning the interface of "human-machine," "human-environment," and "human-human" is applied to explain this case. Safety issues in this report focus on flight crew performance, flight crew decision-making, and pilot fatigue (NTSB, 2002).

Note: O&SHA focus on the problems of operator-machine interface. The risk index (2~4) means it is an "Emergency" case needs immediate response or solution. The risk index (5 ~7) indicates a "Cautious" situation need a fast review and resolution which more information and analysis may be needed to determine the level of risk eroding the entire operational system. The risk index (8 ~10) represents a "Supervisory" case and the reported hazard need a continuous measurement in the future.

International Journal of Applied Aviation Studies, Volume 7, Number 1

The Structure of Safety Climate for Accident Free Flight Crews

Erin E. Block
Edward J. Sabin
Department of Psychology
Saint Louis University
and
Manoj S. Patankar
Department of Aviation Science
Saint Louis University

Shannon Hall, 221 N. Grand Blvd
St. Louis, MO 63103
314-323-2895
erinblock@sbcglobal.net

Abstract

This paper presents a systems-level perspective for understanding and improving safety climate in aviation. Structural equation modeling is used to test a proposed model of flight crew safety in an accident-free commercial aviation company. Analysis of the results indicates two main influences on flight crew safety in the given sample: organizational affiliation and proactive management. Communication, relational supervision, and training effectiveness are also identified as secondary constructs impacting safety for the flight crews in the sample. Applications of the model to future aviation research and use of the model to identify points of leverage for improved safety performance are discussed.

The Structure of Safety Climate for Accident Free Flight Crews

This paper presents a systems-level perspective for understanding and improving safety climate in aviation. A proposed safety climate model is empirically tested using structural equation modeling to analyze data from flight crews working in an accident-free commercial aviation company. This paper responds to the call by various authors (e.g., Cox & Flin, 1998; Li & Harris, 2006; von Thaden, Wiegmann, & Shappell, 2006) to provide scientific evidence that demonstrates the hypothesized relationships between the various components of safety culture and climate.

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

Return to Table of Contents

Cockpit/Crew Resource Management

Major systematic efforts to understand and improve flight safety began in the 1980s with the development of Cockpit Resource Management (CRM). Helmreich, Merritt, and Wilhelm (1999) described the development of five generations of CRM training. Beginning in 1981, CRM used the "Managerial Grid" approach designed by Blake and Mouton (1964) that aimed to improve an individual's competencies in areas such as communication and assertiveness by developing simultaneous concern for coworkers and the task to be accomplished. The "Grid" approach sought to achieve an ideal managerial-style termed "team management" in which excellent work performance was achieved through shared commitment to a common purpose based on interpersonal trust and respect.

The second generation of CRM changed both its name and focus from the cockpit to the crew (Helmreich et al., 1999). *Crew* Resource Management emphasized group dynamics and team functioning in an environment that focused more specifically on aviation operations. In the early 1990s, a third generation of CRM included training on the influence of organizational culture and developed specific behavioral competencies to improve pilot effectiveness. In addition, training was extended to other groups such as flight attendants, dispatchers, and maintenance mechanics.

In fourth generation programs, CRM was purposefully integrated into flight training and became proceduralized by checklists. The fifth (and most recent) generation of CRM attempts to normalize and manage error by recognizing that the error is both inevitable and a valuable source of information. Additionally, Helmreich et al. (1999) argued that CRM provides valuable countermeasures to avoid, trap, and mitigate errors. Additionally, the fifth generation of CRM training attempts to engage all employees in building a robust safety culture and emphasizes cases where errors were detected and managed effectively. This evolution of CRM parallels the development of a four-stage accident theory noted by Wiegmann, Zhang, von Thaden, Sharma, and Gibbons (2004). These four stages include: the technical period, human error period, socio-technical period, and most recently, the organizational culture period.

Organizational and Cultural Influences

Helmreich et al. (1999) and Wiegmann et al. (2004) demonstrated the expanding focus of research to understand error and improve safety. Initially in human factors research, attention was directed to the individual (e.g., pilot error); later research included group/crew level factors, and more recently, studies have encompassed larger cultural/system-level considerations. Each level of analysis offers critical insights, but none is sufficient to understand the complexity of human error by itself. For instance, Hunter (2005) showed the importance of individual factors for aviation safety in his study of hazardous attitudes among pilots. The impact of group dynamics on system safety was demonstrated by the research of Taylor and Thomas (2003), who studied the impact of individual professionalism and interpersonal trust in aviation maintenance groups. They argued that without understanding and building mutual trust between subordinates and managers the effectiveness of safety programs will be thwarted due to cynicism and skepticism

.

Mjos (2004) reported that cultural factors impact the effectiveness of flight crew functioning; in this study a total of 440 pilots were surveyed from three Norwegian commercial airlines using a questionnaire based on Hofstede's (1980) conception of culture. Findings showed that certain cultural factors (e.g., high dominance/ authoritarianism) suppressed communication, which in turn led to more operational mistakes. Soeters and Boer (2000) analyzed a sample of military accidents derived from fourteen NATO air forces; they also used Hofstede's cultural model, and found significant relationships between accident rates and dimensions of individualism, uncertainty avoidance, and power distance.

Von Thaden, Wiegmann, and Shappell (2006) studied the relationship of supervisory and organizational level factors to accidents in commercial aviation. They analyzed 60 accident case reports from the National Transportation Safety Board occurring between 1990 and 2000 that had both pilot error and contributing organizational influences. Analysis of the results indicated that this set of accidents was associated with 10 categories of organizational factors: procedures/directives, training, surveillance, standards, information, supervision, pressure, documentation, substantiation, and facilities.

In general, von Thaden et al. (2006) found that the most common organizational influences related to accidents were inadequate procedures and directives. They noted that, as airlines increased in size, "organizational problems appear to shift from issues of training and surveillance to issues of information sharing, communication, and documentation" (p. 253). Von Thaden et al. concluded that instead of blaming the pilot for accidents, aviation would be better served by correcting supervisory and organizational problems that would improve the overall system.

Li and Harris (2006) reached a similar conclusion on the impact of supervisory and organizational issues on aviation. They studied 523 accidents occurring between 1978 and 2002 in the Republic of China Air Force. Findings indicated relationships between organizational inadequacies and pilot error. They noted chains of events beginning with poor decisions made by upper levels of command, which in turn impacted supervisory practice, which set the stage for impairing pilot performance that resulted in accidents.

Safety Culture

Cox and Flin (1998) noted that use of the term "safety culture" began with the investigation of the explosion at the Chernobyl nuclear power plant in 1986, and Meshkati (1997) traced the prominence of the "safety culture" concept in aviation to the NTSB investigation of the crash of Continental Express Flight 2574 in September 1991. Wiegmann et al. (2004) provided a review of the substantial literature that has been generated around safety culture and climate during the past two decades. These authors commented on the diverse range of definitions for safety culture that have been developed by high consequence industries. From these definitions, they distilled a set of seven features common to the various articles on safety culture. According to Wiegmann et al. (2004), safety culture: operates at the group/system level; concerns formal safety issues; involves all organizational members; impacts work behavior; reflects the relationship between rewards and safety performance; views errors/accidents as valuable sources of information for learning; and is relatively stable. Based on their analysis of the safety culture literature Wiegmann et al. conclude that there are five primary global indicators of safety culture that include the following: 1) organizational commitment; 2) management involvement, 3) employee empowerment; 4) reward systems, and 5) reporting systems.

Gibbons, von Thaden, and Wiegmann (2006) described results of a test for the five-factor model of safety culture proposed by Wiegmann et al. (2004). They analyzed 503 surveys from pilots and managers of a large U.S. commercial airline via confirmatory factor analyses, and determined a four-factor model (four main factors and eleven subscales) provided a more accurate representation of safety culture than their initial five-factor model. The four primary factors in the Gibbons et al. model are as follows:

1) a formal safety system (e.g., reporting system),
2) an informal safety system (e.g., professionalism),
3) operations personnel (e.g., chief pilots), and
4) organizational commitment (e.g., safety values).

Safety Culture vs. Safety Climate

Even the briefest review of literature reveals any number of definitions of "safety culture," depending on such things as the author(s) research perspectives or the industry being studied (Wiegmann et al. 2004). In fact, Wiegmann et al. provided a table of 13 different definitions of safety culture from various sources (p. 122-123). As mentioned previously, Wiegmann et al. distilled seven critical features common to these definitions; a key feature among these seven, "safety culture is relatively enduring, stable, and resistant to change" (p. 123). This was contrasted with the common features of "safety climate" definitions also identified by these authors. Safety climate definitions include the following: 1) perceptions of the state of safety at a particular time, 2) closely concerned with intangible issues such as situational and environmental factors, and 3) a "snapshot" of safety culture, relatively unstable and subject to change. This notion of climate as a temporary image of the underlying safety culture is echoed by Schein (1992), who wrote, "climate will be a reflection and manifestation of cultural assumptions" (p. 230).

While there has not always been a great deal of agreement in the literature with regard to the specifics of these definitions, and in fact some consider the two terms interchangeable, the unique features of each as identified by Wiegmann et al. (2004) provided a solid foundation for understanding safety climate and safety culture as separate concepts. Based on this foundation, the present study focuses on safety climate within the flight crews of a commercial aviation organization. The survey items used assess perceptions of organization members, often of intangible facets of the organization's functioning, and this, connected with the intangible nature of the outcome items of the survey, means a temporal snapshot of safety in this organization has been captured here. This does not, however, diminish the value of the model here presented. Rather, it provides a foundation for future research for understanding key factors that may be present and foster a climate of safety. Understanding how these climate factors tie into underlying cultural factors would, in the future, provide an even more comprehensive picture of the nature of safety in the aviation setting.

The Purpose-Alignment-Control (PAC) Model

Building on existing literature, Patankar, Bigda-Peyton, Sabin, Brown, and Kelly (2005) proposed a systemic model of safety culture that identified three primary factors: purpose, alignment, and control (PAC Model). Each of the three factors consists of other latent constructs found in the aviation literature, e.g., institutional identity, leadership, relationships, and goal sharing. An important aspect of the PAC model is its integrated nature; while there are three main factors present in the model, they are interdependent with each other, and describe the multiple influences on safety culture as part of a single system. This strategy of thinking of safety culture not as an isolated part of the aviation environment, but as a construct completely integrated with and relevant to all aspects of the environment, is key to understanding if the appropriate leverage points for improvement are to be identified.

In the PAC model, purpose is defined as understanding, accepting and focusing on the mission and key goals of the organization. An organization, with a positive safety culture, values and unequivocally communicates safety as a critical part of its core corporate identity. Alignment involves the systematic coordination of technical, organizational, team, and individual resources to fulfill mission critical goals. Alignment results in high consistency between words and actions. In a positive safety culture, leaders align human, technical, and cultural assets to produce effective organizational and team structures. In addition, these assets produce effective communication networks, technical infrastructures, professional accountabilities, training opportunities, and proactive management systems in a social environment that fosters individual development, interpersonal trust, and mutual respect.

Control involves a set of key beliefs that individual and team actions can make a critical difference in recognizing, mitigating, or preventing risks, errors, and hazards in complex socio-technical systems. Control in a positive safety culture centers on the notions of individual initiative and collective efficacy to detect and correct error. This PAC concept of control is related to findings on locus of control (e.g., Rotter, 1966, 1990) and self-efficacy (e.g., Bandura, 1997) that have proven to have powerful impacts on human behavior; specific to aviation research. Hunter (2002) has developed a scale that measures locus of control as it relates to aviation safety. Research in aviation has also studied the relationship of self-efficacy to flight training (Davis, Fedor, Parsons, & Herold, 2000) and pilot-automation complacency (Prinzel, 2002).

The PAC model proposes a number of relationships among various constructs related to aviation safety, relationships that to this point have not been empirically evaluated. To evaluate complex relationships of this type, structural equation modeling (SEM) is a powerful multivariate statistical tool (Schumacker & Lomax, 2004). Previously in aviation research SEM has been used to study the influence of mission complexity on mental workload, situational awareness, and pilot performance during simulated missions with active fighter pilots (Svensson & Wilson, 2002). In the present study, SEM is employed to empirically test multiple relationships regarding safety climate in flight crews, relationships that are proposed in the PAC model of safety culture. The present study seeks to use the data set from Patankar (2003) to create a more complete picture of the variables influencing safety outcomes in flight crews, as well as establish an empirical link between the existing Patankar (2003) data and the PAC model theorized by Patankar et al. (2005). Identifying links between perceptions of safety (safety climate) in flight crews and the underlying notions of culture presented in the PAC model is important in developing a comprehensive understanding of the variables that may influence safety in a high-consequence industry (such as aviation). The fit of this flight crew safety model to the data is evaluated, links to the PAC model are discussed, future research options and the utility of this model for identifying points of leverage to enhance aviation safety are explored.

Method

Dataset

Participants were members of flight crews (captain, first officer, or second officer/engineer) employed by a large commercial aviation company in the United States. The current dataset consisted of 281 flight crewmembers, 95.7% male, who completed the *Organizational Safety Culture Questionnaire.* The average age of the respondents was 41.37 (s = 9.08) years, with a mean service time of 7.87 (s = 5.21) years as a pilot for this company.

Procedure

The data set for the present study is derived from Patankar (2003), who studied a large commercial aviation company that had an accident-free safety record for over 20 years. The fifty-item Organizational Safety Culture Questionnaire (OSCQ, used by Patankar, 2003) was completed by 399 employees (55% return rate), including: flight crews, maintenance crews, and other aviation support staff.

Patankar used a factor analytic technique to identify eight key constructs present in the entire data set (N = 399): pride in company, professionalism, safety, supervisor trust and safety, personal stress, assertiveness, safety compliance, and hazard communication. Significant differences were found between these groups for measures of: pride in company, safety opinions, and supervisor trust. Due to these significant differences between professional groups, the present study focused only on the flight crew group, which included questionnaires completed by 281 pilots (56% return rate). Additionally, Patankar et al. (2005) theorized the “Purpose-Alignment-Control (PAC)” model of safety culture. This model hypothesized that factors such as institutional identity, information flow, and leadership are relevant for influencing safety, specifically in an aviation setting.

In order to integrate the flight crew portion of the *Organizational Safety Culture Questionnaire* (*OSCQ)* data with the theorized PAC model, the authors employed an item-mapping technique. Two raters familiar with both the PAC model and the *OSCQ* dataset first performed an exploratory factor analysis on the *OSCQ* data to identify item distribution compared to the factors discussed in Patankar et al. (2005). The raters then took these items and their related factors and mapped them to the theoretical constructs in the PAC model, by comparing the existing *OSCQ* items to potential items presented in the PAC model. The raters found a number of items and factors that were similar for both the *OSCQ* and the PAC model. Any differences between the raters were discussed and referred to a third rater for additional input until agreement on the constructs and their item indicators was reached. These were then the items and constructs selected for testing. Confirmatory factor analysis was used to support the mapping technique that the theoretical constructs selected for the structural model had the appropriate item indicators mapped to them. In the present study, of specific interest was the desire to go beyond identification of safety-relevant factors in order to understand the relative weight of influence these factors may have and the ways these factors may interact with each other to influence safety. In addition, the authors sought to improve understanding of how safety culture variables proposed in the PAC model may compare to climate factors present among flight crewmembers. In order to accomplish this, a structural equation modeling procedure was selected to analyze the flight crew portion of the *OSCQ* dataset.

Structural equation modeling (SEM) is a statistical analysis tool that tests various theoretical models that "hypothesize how sets of variables define constructs and how these constructs are related to each other (Schumacker & Lomax, 2004, p. 2)." The goal of SEM analysis is to determine the extent to which a given theoretical model is supported by data collected from a relevant sample. Because the item-mapping technique employed by the authors necessarily created a degree of ambiguity in the ways in which the different constructs may influence safety outcomes (i.e., whether a construct had a direct or indirect effect on safety was uncertain from an *a priori* theoretical perspective), an alternative theoretical model was tested using SEM procedures to find the model that best fit the existing data set. This alternative model tested whether each latent construct had a direct effect on safety outcomes (Appendix A). According to Hoyle (1995), this inductive approach is acceptable from an analytic perspective if fit indices demonstrating the authors' reasoning in selecting their "best-fitting" model are also provided. In the present study, selection of the best-fitting model is evidenced most clearly in the value of the path coefficients between the theoretical constructs (shown in Appendix A).

Results

As described above, two models of flight crew safety were tested using SEM analyses, and were compared based both on fit indices as well as on the strength and direction (positive or negative) of the path coefficients between each latent construct to determine the appropriateness of the model. When using SEM, judging the fit of a model to a given dataset through use of a single index is not sufficient (Vandenberg & Scarpello, 1990), as each fit index has various strengths and weaknesses associated with it. Judgment of fit relies on converging evidence from different indicators provided by the statistical analysis. Fit indices are used to indicate the hypothesized model's fit to the sample data. For the present study, four indicators of model fit were used: path coefficients, the root mean square error of approximation (RMSEA), the comparative fit index, and the Tucker-Lewis index. The selection of each fit indicator is described below.

Results from the path analysis indicated the final theoretical model selected by the authors (Figure 1) demonstrated good overall fit with the data. The standardized path coefficients describe the relationships among the latent constructs and the strength of that relationship with regard to the dependent (outcome) variables. As can be seen in Figure 1, these coefficients (also called *betas)* were all significant at the p < .05 level. The coefficients and fit indices for the alternative theoretical model tested can be found in Appendix A. The root mean square error of approximation (RMSEA) is a global measure of model fit; for research in the social sciences, an RMSEA value less than 0.08 is considered acceptable fit. For the flight crew sample in the present study, RMSEA = 0.076.

The comparative fit index (CFI) compares the authors' hypothesized model with a null model (one in which all latent constructs are assumed to be uncorrelated). CFI greater than 0.90 indicates acceptable model fit, for the present study CFI = 0.96. The Tucker-Lewis index (TLI, also called the non-normed fit index) is used to compare alternative models or to compare a single model to a null model; TLI scores greater than 0.90 are acceptable, and greater than 0.95 indicate the model is a good fit with the data. The Tucker-Lewis index for the final model was 0.96. McDonald and Marsh (1990; cited in Schumacker & Lomax, 2004) concluded the TLI was one of only two fit indices to remain unbiased in finite samples, and recommended using it for testing null or alternative models. Given the size of the data set for the flight crews (sufficient for SEM analysis, but small compared to the general SEM literature), the TLI is an appropriate fit index to report.

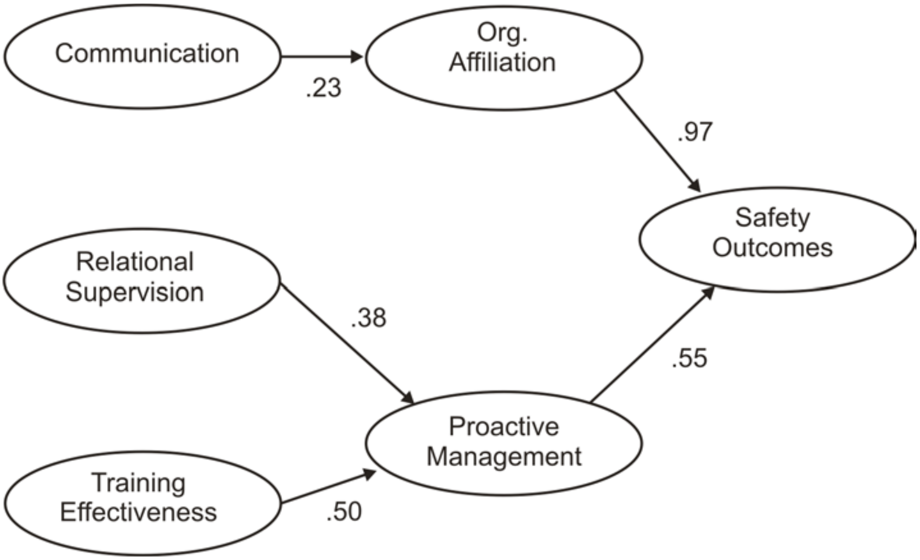


Figure 1. Safety Outcomes Model for Flight Crews

As can be seen in Table 1, the items selected to represent each construct appear to be appropriate indicators for the latent variables. While .70 is often considered the cutoff point (i.e., values higher than .70 are desired) in common discussion of structural models, the authors felt the unique nature of the applied sample and the use of the item-mapping procedure called for a degree of leniency with regard to the strength of the lambda coefficients; lambda referring to the value of the relationships between the observed variables and the latent constructs (lambda values are referred to as *factor loadings).* Thus, the authors established a .50 cutoff point for items – any item with a lambda below .50 was not included in the final model. Future tests of the model may include creating survey items to more accurately assess the latent variables and improve the lambda coefficients.

Table 1
Estimates for Measurement Parameters in Figure 1

Safety Outcomes	λ
37. My company is the "Best in the Business."	.65
43. Safety in this organization is largely due to our collective commitment to safety.	.56
Organizational Affiliation	
1. Our company's safety practices are consistent with the published corporate values and mission.	.63
26. Working here is like being part of a large family.	.64
27. I am proud to work for this company.	.81
Communication	
21. All employees should make the effort to foster open, honest, and sincere communication.	.90
23. A debriefing and critique of procedures and decisions after a significant task is completed is an important part of developing and maintaining effective coordination.	.70
25. My coworkers valued consistency between words and actions.	.54
40. Communication between pilots and mechanics should be encouraged.	.54
Proactive Management	
28. My organization closely monitors and corrects any deviations from established quality standards.	.79
31. We are proactive in hazard identification and management.	.79
42. Safety in this organization is largely due to adherence to the standard operating procedures.	.62
45. Safety in this organization is largely due to positive changes resulting from our past experience with incidents and/or accidents.	.56
Training Effectiveness	
14. We receive an adequate amount of safety training.	.65
18. I know the proper channels to route questions regarding safety practices.	.61
32. I am aware of a self-reporting system for accidents.	.54
38. I am adequately trained to conduct all of my job duties and responsibilities.	.67
Relational Supervision	
3. I feel comfortable going to my supervisor's office to discuss problems or operational issues.	.84
22. My supervisor can be trusted.	.91
33. My supervisor listens to me and cares about my concerns.	.91
35. My supervisor protects confidential or sensitive information.	.75
48. I trust my superiors to choose safety over performance.	.69
50. My suggestions about safety would be acted upon if I expressed them to my supervisor.	.77

The final theoretical model shows two primary drivers of safety outcomes: *organizational affiliation* (similar to "pride in company" from Patankar [2003]) and the organizational identity and relationships factors from Patankar et al. [2005]) and *proactive management* (items partially derived from the "safety opinions" factor from Patankar (2003), as well as items identified through the factor analysis). Organizational affiliation is, in turn, influenced by *communication*; that is, honest communication among employees. These three constructs (communication, mediated by affiliation, leading to perceived safety outcomes) linked together can be compared to the “Purpose” portion of the PAC model, which Patankar et al. describe as an ability to “collapse constraints, self-organize, and sharply focus on the goal (p. 46).” When positive communication patterns among employees are occurring (or perceived by employees to be occurring), this may affect the extent to which employees feel affiliated in a personal way with the organization. Both communication and organizational affiliation depend on this notion of “self-organizing and sharply focusing on the goal” as development of either one in an organizational setting cannot be mandated by leadership, but must develop. The communication – affiliation link--may be present in the organization’s safety climate and serve as the indicators of underlying “purpose” in the organization’s culture. Safety outcomes are affected by this sense of “purpose,” this communication – organizational affiliation link. The model suggests that in the presence of open communication and a positive personal connection with the organization, positive perceptions of the safety of the organization and the collective commitment of the entire organization to safety also improve.

The second portion of the PAC model is *alignment* – aligning individual and organizational resources to accomplish a goal. In the structural model for the flight crews presented here, alignment is identified with the presence of the *relational supervision* and *training effectiveness* constructs. Relational supervision is the way in which employees view their immediate supervisor as a person who is trustworthy and concerned about safety at an individual as well as an organizational level. Training effectiveness describes the resources available in the organization for successful job accomplishment: awareness of a safety reporting system, adequate training, and appropriate channels for safety-related questions. When both of these constructs influence *proactive management* in a positive way, there is *control* (the third portion of the PAC model) over safety outcomes.

Proactive management describes employees’ perceptions of the organization as being concerned with the preemptive identification and correction of hazards, deviation from quality standards, and the willingness to make changes based on past errors or incidences. These perceptions mediate the influence of relational supervision and training effectiveness on safety outcomes. Without the presence of this organizational management, the positive perceptions of supervisors and the perceived effectiveness of training for employees are far less likely to influence safety in an aviation organization. Patankar et al. (2005) define control in terms of outcomes; for example, pilots “try to prevent an impending accident because they believe that it is preventable (p. 46).” Similarly, the flight crew model presented here indicates that the presence of the ‘alignment constructs’ of relational supervision and training effectiveness may not influence safety without the presence of proactive management to create an organizational environment in which problems and safety hazards are preventable.

The notion of control described in the PAC model and identified by the constructs in the flight crew safety model is also similar to that of *self-efficacy*, described in the psychological research literature as “personal judgments of one’s capabilities to organize and execute courses of action to attain designated goals”

(Zimmerman, 2000, p. 83). According to the self-efficacy literature, individuals' decisions to work actively toward a particular goal depend in part on their internal determination that they are capable of successfully attaining said goal – i.e., Patankar et al.'s statement regarding pilots working to prevent accidents they believe are preventable. For flight crewmembers, the safety model suggests a proactive management structure contributes to the creation of organizational situations in which the members' self-efficacy is increased. Increased self-efficacy may then lead flight crews to work to prevent incidents or to enhance existing safety outcomes because they believe it is in their power to do so, and that the organization is supportive of such efforts.

Discussion

The present study sought to increase understanding of the ways in which various factors may influence safety outcomes for flight crewmembers. The intention being, the creation of a model that suggests points of intervention to positively affect aviation organizations; that is, places or ways to assist the organization in making change such that it will be more likely to positively influence safety outcomes. Secondly, it was hoped, this research would create a link between a model for a flight crew safety climate and the PAC model theorized by Patankar et al. (2005), which indicated key parameters for safety culture in aviation. The SEM analyses conducted on the data collected using the *Organizational Safety Culture Questionnaire* provided support for this link and suggested specific theoretical constructs as potential points of intervention for an aviation organization interested in improving safety outcomes.

The structure of both the PAC model and the flight crew safety model overlap with existing psychological and organizational literature, lending further credence to continued research on these models to foster understanding of relevant safety factors. Bandura's work on self-efficacy in individuals (1997, as well as others) and research on theories of individual locus of control (e.g., Rotter, 1990), indicate the "control" aspect of the PAC model (as well as the "proactive management" portion of the flight crew safety model) is tapping into internal processes of these organization members. Similarly, the notion of alignment as one of the drivers of organizational performance has also been discussed in the literature (Middleton & Harper, 2004; Semler, 1997). The flight crew safety model presented here taps into this research in the links with relational supervision and training effectiveness and finds support for it in this aviation setting.

While the link between communication and organizational affiliation in the flight crew safety model may not appear as strong as some of the other relationships in the model, this may have two partial causes: 1) the use of a pre-existing dataset collected by a survey not designed with regard to the flight crew safety model (and thus not containing the most salient communication items), and 2) the high link between organizational affiliation and safety outcomes may be masking additional influence from the communication construct. Again, this may be a factor of the use of existing items to test a new model of flight crew safety. The overall model demonstrated acceptable fit, though, and the authors believe future research on the flight crew safety model will lead to refinement and validation of its general structure, including the link between communication and organizational affiliation.

As can be seen in Appendix A, the direct-relationship model, also tested, exhibits comparable fit indices, suggesting it is also a viable model of flight crew safety. This may again partially be a function of the use of a pre-existing dataset. What then becomes important is looking not only at fit indices, but also at the path coefficients between each latent construct in the two models. Standardized path coefficients that are not significant indicate a lack of a meaningful relationship between the two constructs, while path coefficients exceeding 1.00 (or -1.00) may instead indicate an issue of multicollinearity (i.e., other latent constructs may overlap in the nature of the variance for which they are accounting).

The strength and direction of the path coefficients had to make sense from a theoretical perspective; with the alternative model, it was clear from the direction of the path coefficient that the model was not a correct interpretation of the data (e.g., in the direct-relationship alternative model, the path coefficient between Relational Supervision and Safety Outcomes is -.82, suggesting low relational supervision would drive higher or more positive safety outcomes, which is not supported by other aviation safety research or theoretical discussion). It is hoped, though, that future research may provide data that demonstrate the relationships presented in the flight crew safety model with a greater degree of clarity, and even more clearly separates the flight crew safety model from the alternative model tested.

The flight crew safety model identifies several potential applications for intervention. Organizational Affiliation among employees can be a powerful tool to enhance safety; perhaps if employees feel this personal relationship to the aviation organization, it may strengthen their desire to ensure the best possible safety conditions are present. Proactive Management can be improved in a number of ways to yield positive improvement in safety; for example, introducing systems that enable all employees to aid not only in hazard identification, but also in hazard repair – so that employees are both part of problem finding as well as problem solving. An aviation organization that establishes ways to learn from past experiences, both positive and negative (an example would be the after action reviews used in the military) could directly improve safety outcomes.

Organizational Affiliation and Proactive Management are two drivers that cannot function effectively without the presence of communication, relational supervision, and effective training for employees at all levels. An analysis of an aviation organization with the intent to improve safety would gain maximum benefit by understanding that these more indirect constructs must be in place and used by employees in order for improvements to affiliation or proactive management to ultimately influence safety.

The present study utilized existing data to create an initial model of flight crew safety. Unfortunately, the nature of the sample (i.e., pre-existing data, limited sample size) meant the flight crew safety model could not be adequately tested and retested by splitting the sample or collecting additional data. In addition, the Organizational Safety Culture Questionnaire (OSCQ) has not been thoroughly validated or checked for reliability; however, because the OSCQ is a combination of relevant questions from several other, previously established questionnaires (Patankar 2003), the authors felt the data gathered from its administration could be used with a degree of confidence in the present study. Future research would be beneficial to confirm this assumption. Additional research is needed to confirm the model and the relations between the various constructs in it.

Based on this model, future surveys of flight crews may be revised to incorporate items relating to the constructs identified here. This may in turn provide the aviation organizations with more specific information on the status of these constructs as well as on the status of identified safety outcomes. Additional data collection would also be beneficial for the refinement and validation of the flight crew safety model presented here; since the present study relied upon a pre-existing dataset, the flight crew model is necessarily somewhat tentative. It is hoped publication of the model will provide an impetus for further research to refine the model and even increase its ability to identify safety-related constructs. While the model presented here was designed for flight crews, future research that can replicate these results as well as extend them to other populations (especially other aviation populations), would be desirable.

Understanding the relationship between safety climate and underlying safety culture is an important step in improving safety. By linking a model of safety climate to the PAC model of safety culture, the authors hope to foster additional research on both of these constructs. The flight crew safety model represents a step forward in understanding the relationships among some of the key constructs influencing safety outcomes in an aviation setting; the authors hope the present study's work to improve this understanding in turn fosters an increased ability to act in aviation organizations to improve safety.

References

Bandura, A. (1997). *Self-efficacy: the exercise of control.* New York: Freeman.

Blake, R.R., & Mouton, J.S. (1964). *The Managerial Grid*. Houston: Gulf.

Cox, S., & Flin, R. (1998). Safety culture: Philosopher's stone or man of straw? *Work and Stress*, *12*(3), 189-201.

Davis, W.D., Fedor, D.B., Parsons, C.K., & Herold, D.M. (2000). The development of self-efficacy during aviation training. *Journal of Organizational Behavior, 21(8)*, 857-871.

Gibbons, A. M., von Thaden, T.L., & Wiegmann, D. A. (2006). Development and initial validation of a survey for assessing safety culture within commercial flight operations. *The International Journal of Aviation Psychology*, *16*(2), 215-238.

Helmreich, R.L., Merritt, A.C., & Wilhelm, J.A. (1999). The evolution of crew resourcemanagement training in commercial aviation. *The International Journal of Aviation Psychology*, *9*(1), 19-32.

Hofstede, G. H. (1980). *Culture's consequences: International differences in work-related values.* Beverly Hills, Sage.

Hoyle, R. H. (1995). *Structural equation modeling: concepts, issues, and applications.* Thousand Oaks, CA: Sage Publications Inc.

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

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

Li, W-C, & Harris D. (2006). Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. *Aviation, Space and Environmental Medicine*, *77*(10), 1056-1061.

Meshkati, N. (1997). Human performance, organizational factors and safety culture. Paper presented at NTSB National Summit on Transportation Safety. Washington, DC.

Middleton, P. & Harper, K. (2004). Organizational alignment: A precondition for information systems success? *Journal of Change Management, 4*(4)*,* 327-338.

Mjos, K. (2004). Basic cultural elements affecting the team function on the flight deck.*The International Journal of Aviation Psychology*, *14*(2), 151-169.

Patankar, M.S., (2003). A study of safety culture at an aviation organization. *International Journal of Applied Aviation Studies*, *3*(2), 243-258.

Patankar, M.S., Bigda-Peyton, T., Sabin, E., Brown, J., & Kelly, T. (2005).*A comparative review of safety cultures.* Federal Aviation Administration: ATO-P Human Factors Research and Engineering Division. (Grant No. 5-G-009).

Prinzel, L.J. (2002). The relationship of self-efficacy and complacency in pilot- automation interaction. NASA (CASI): Hanover, MD. NASA/TM-2002-211925.

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

Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *The American Psychologist, 45*(4)*,* 489-493.

Schein, E.H. (1992). *Organizational Culture and Leadership.* (2nd ed.). New York: Wiley& Sons, Jossey-Bass; Third edition, 2004.

Schumacker, R.E. and Lomax, R.G. (2004). *A Beginner's Guide to Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum.

Semler, S. W. (1997). Systematic agreement: A theory of organizational alignment. *Human Resource Development Quarterly, 8*(1)*,* 23-40.

Soeters, J.L., & Boer, P.C. (2000). Culture and flight safety in military aviation. *TheInternational Journal of Aviation Psychology, 10*(2)*,* 111-133.

Svensson, E.A. I., & Wilson, G.F. (2002). Psychological and psychophysiological models of pilot performance for systems development and mission evaluation.*International Journal of Aviation Psychology, 12*(1), 95-110.

Taylor, J.C., & Thomas, R.L. (2003). Toward measuring safety culture in aviationmaintenance: The structure of trust and professionalism. *The International Journal of Aviation Psychology, 13*(4), 321-343.

Vandenberg, R. J., & Scarpello, V. (1990), The matching model: An examination of the processes underlying realistic job previews, *Journal of Applied Psychology*, *75*, 60-67

Von Thaden, T.L., Wiegmann, D. A., & Shappell, S. A. (2006). Organizational factors incommercial aviation accidents. *The International Journal of Aviation Psychology, 16*(3), 239-261.

Wiegmann, D.A., Zhang, H., von Thaden, T. L., Sharma, G., & Gibbons, A.M. (2004).
Safety culture: An integrative review. *The International Journal of Aviation Psychology, 14*(2), 117-134.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary Educational Psychology, 25(1), 82-91.

International Journal of Applied Aviation Studies, Volume 7, Number 1

Testing a Nonveridical Aircraft Collision Avoidance System: Experiment 1

William R. Knecht
Maneuver Space Technologies
Oklahoma City, OK USA

Abstract

Veridical displays represent realistic scenes. However, they are limited when displaying certain types of aircraft maneuver information. This report describes the first major formal, successful test of 4CAS, a nonveridical, 4D aircraft collision avoidance system defined by three coordinate axes of aircraft heading, speed, and altitude. In Experiment 1, eight licensed general aviation pilots each flew eight simulated free-flight scenarios with the goal of deviating as little as possible from a pre-assigned course, while still maintaining standard enroute separation from traffic. Half the scenarios were flown with only a cockpit display of traffic information (CDTI) to assist separation maintenance. The remaining half was flown with the CDTI plus 4CAS. The CDTI+4CAS scenarios showed performance superiority over the baseline CDTI-only scenarios for one dependent measure of maneuver efficiency, two measures of maneuver safety, and two measures of user workload. Results suggested that nonveridical displays might be useful in aircraft separation maintenance.

Introduction

Currently, most U.S. commercial aircraft do not fly efficiently from point-to-point, but follow segmented jet ways in en route airspace (en route is the “long-haul” airspace, more than 40 nm distant from departure or destination). As Figure 1 shows, these segmented jet ways effectively add unnecessary travel distance. In 2005, U.S. airlines spent over $36B on fuel alone (BTS, 2006). Enabling aircraft to fly directly from point of departure to destination would save time, lower aircraft component stress, and lower fuel use and upper-atmospheric $C0_2$ deposition by approximately 6% (ORA, 1998).

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

Return to Table of Contents

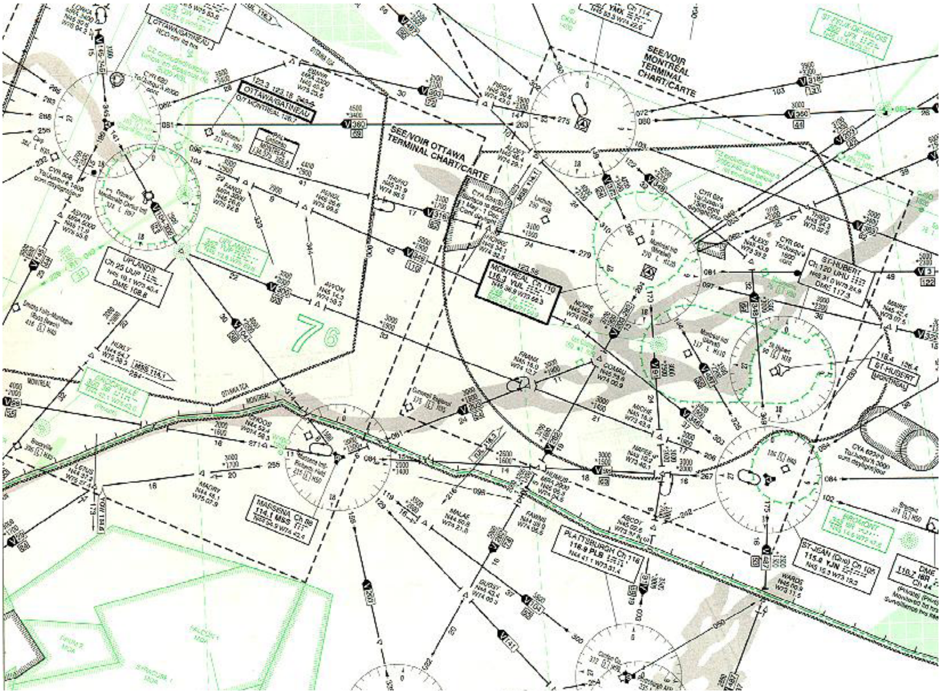


Figure 1. An aeronautical chart showing high-altitude enroute jet ways. These routes are not direct, and add unnecessary distance to most commercial flights.

Addressing these concerns, U.S. air traffic controllers are starting to transition to direct routing. However, direct routing increases airspace complexity as traffic paths begin crossing in new and more complex ways. Flight efficiency is gained at the expense of increased aircraft maneuver frequency and complexity. Ironically, direct routing represents a gain for all National Airspace System (NAS) stakeholders except the very pilots and air traffic controllers who key line-level experts are making the NAS work.

To some degree, well-designed aircraft separation maintenance/collision avoidance technology should be able to mitigate these safety and workload concerns, while still allowing the desired efficiency gains. This is an empirical issue, and the subject of considerable research. Krozel (2000) reviewed much of the literature.

Managing the NAS involves triple strategic goals of navigational safety, efficiency, and ease-of-use. The technology to be described in this paper is not the only one capable of accomplishing these goals. Prototypes of mathematically elegant, precise, fully automatic collision avoidance systems (auto-CAS) also exist to address the situation (Eby & Kelley, 1999; Hoekstra, Ruigrok, & van Gent, 2000; Johnson, Bilimoria, Thomas, Lee, & Battiste, 2003). However, a precision *manually initiated* technology (MI-CAS) is also appropriate because political and bureaucratic exigencies prevent auto-CAS from being implemented. From the operational perspective, no technology yet devised is 100% reliable. Moreover, from the psychological and political perspectives, pilots, air traffic controllers, and passengers can be expected to resist technology that threatens jobs or safety. Consequently, having a human in the loop will always be a point of concern for expert operators, consumers, NAS stakeholders, engineers, and human factors researchers.

The current research addresses these issues. The system under scrutiny is called 4CAS (*4-Dimensional Collision Avoidance System*). The research intent is to demonstrate that safe, easy, efficient direct routing of aircraft is possible with a combination of a manual CAS plus a CDTI. It is hoped that this combination may someday rival auto-CAS in performance, yet also be capable of complementing auto-CAS by providing a way to display the intent of auto-CAS-generated maneuvers. A 4CAS prototype exists. Does it allow safe, efficient navigation through extremely crowded airspaces with relatively little operator training and effort? If so, does it do so reliably and significantly better than currently fielded MI-CAS technology?

Introductory concepts underlying the technology

Air traffic control (ATC) is always challenging. Direct-routing is even more so. The most difficult case is "pure free flight"—a theoretical system of direct routing where pilots themselves would control separation maintenance with no help at all from ground-based ATC (RTCA, 1995). Currently, there is no free flight in U.S. commercial air travel. Nevertheless, because free flight represents the "complex" end of a complexity/workload continuum, it is often used by planners and engineers to design upgradeable traffic control systems, systems capable of handling increased traffic and anticipated changes in procedures and technology.

To illustrate the kind of complexity we might encounter in free flight, consider the following purely hypothetical traffic situation (Figure 2). Let the green path depict *O*, the pilot's own ship (own ship), a Boeing 737-400, along with two intruder ships *I1, I2* (white paths). Assume all aircraft are in straight-and-level flight at FL320 (32,000 ft), all cruising at 290 kt IAS (.78 mach). This is a moderately difficult situation requiring maneuver by the own ship.

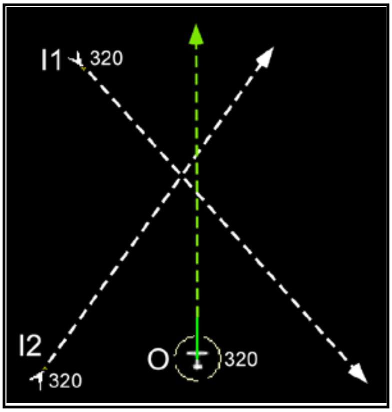


Figure 2. Hypothetical free flight situation requiring maneuver.

So, why is this situation difficult? Visualize the task at the level of information representation. Right now, Figure 2 depicts a map. Maps are veridical. *Veridical* means, “coinciding with reality.” Maps coincide with spatial reality. Veridicality implies representation of extension in physical space—width, depth, and height.

However, the fact is that humans have not evolved to interact effortlessly with maps. The perceptual tasks we do quickly and easily are recognizing faces, walking through the world, manipulating 3D objects, and so forth (Gibson, 1979). Even though each of those ecological tasks is actually quite complex, they arguably appear natural to us because countless generations of our ancestors were differentially selected for just such real-world abilities.

The same is not true of solving aircraft maneuver problems on flat-screen map displays. That particular task had no analogue common to the world of our ancestors. Consequently, there was no particular need for a nervous system capable of excelling at it. Moreover, this is almost certainly the phylogenetic “reason” why air traffic control on map displays seems difficult to us.

Given the phylogeny, the point is that *if* the veridical task of aircraft maneuver could somehow be *transformed* into the kind of task people do easily, then solving even difficult traffic problems should become easier. The next step then becomes how to do such a transformation.

For data visualization in chemistry and phy*sics*, *non*veridical display is commonly used. Perhaps this is applicable to the problem at hand.

Consider the example of a chemical process that is dependent on three factors—temperature, pressure, and catalyst concentration. These factors can be represented by numbers, but are not veridical because they are not width, depth, or height. We can use a 3D coordinate system to represent a *state space* (Figure 3). This state space will have one axis representing temperature, a second representing pressure, and a third representing catalyst concentration. Inside this space, we now represent many triple combinations of temperature, pressure, and catalyst, stacked like tiny glass blocks inside a box.

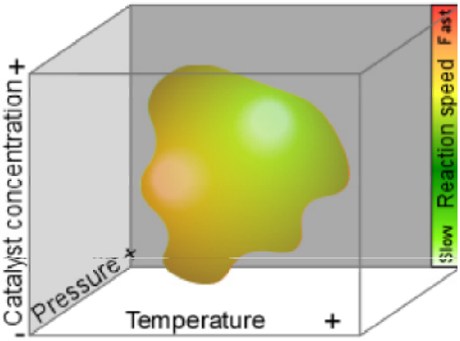


Figure 3. A state space dimensionalized by temperature, pressure, and catalyst concentration. Color depicts reaction speed.

In Figure 3, red = “fast,” yellow = “moderately fast, green = “moderately slow,” white = “slow,” and transparent = “no reaction.” Moreover, since a “dimension” can represent *any* quantity that varies, if we color-code each small block to represent reaction speed at that particular combination of variables, this increases the overall space dimensionality to 4D.

The potential benefit to creating such a 4D state space is that it takes a complex set of abstract physical interactions and transforms it into a much simpler, more concrete “world” that looks a lot more like what we see every day. This new world arguably appeals far more to our perceptual and cognitive strengths than, say, looking at a 3D table of numbers. Such a table would contain essentially the same information, but might look like an impossibly cluttered, meaningless jumble.

A “maneuver space” that represents maneuver information

We can see how state spaces simplify information display. Now, consider briefly what aircraft “maneuver” is. We can turn left or right. We can go faster or slower. We can climb or descend. At any given instant in time, the exact values of these three quantities constitute the three basic “dimensions” of an aircraft’s maneuver “state.”

Multidimensional state spaces are not typically applied to aircraft separation. However, given how we just defined maneuver, we can certainly construct a state space based on our aircraft autopilot’s *heading*, *speed*, and *altitude*.

Figure 4. Our autopilot, showing its three “dimensions” of *heading*, *indicated airspeed*, and *altitude*. Each triplet can comprise one “maneuver” inside a state space.

Figure 5 shows a slightly tilted view of this new type of space, with a horizontal axis showing own ship heading, an in-out axis for indicated airspeed, and a vertical axis for altitude. Inside this space, we can now represent many triple combinations of *maneuver*, stacked like tiny glass blocks inside a box.

This entire state space can be called a *maneuver space* (MS) because it is “a space constructed of maneuvers” (Knecht & Smith, 2000). Since only one dimension is truly veridical (altitude), the entire space is classified as nonveridical. To understand MS at the highest conceptual level, picture it as a “**maneuver-hypothesis tester**.” It represents a series of hypothetical maneuvers, each tested for safety. Specifically, for each possible maneuver, *“If I were flying straight and level at that heading* H*, speed* S*, and altitude* A*—would I be in conflict with any obstacle in the near future? If so, how much time would I have to maneuver away safely?”*

In practice, the actual conflict prediction is done by a hardware/software *conflict probe* (Kuchar & Yang, 2000). The probe calculates “safe separation” by normal enroute separation standards (5 nm lateral/1000’ vertical separation). Although the probe can only reliably look a short time into the future (6-10 minutes), experience shows that this is adequate time to maneuver safely in virtually all realistic traffic situations (Knecht & Hancock, 1999).

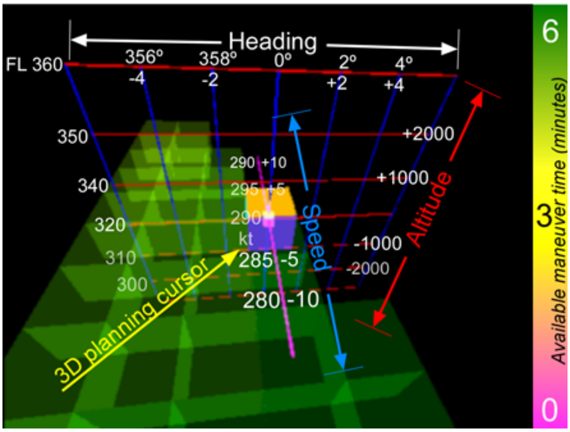


Figure 5. A slightly downward-looking view of how Figure 2's veridical space transforms into nonveridical maneuver space (MS).

Figure 5 embodies eight essential operational definitions:

1) Each cube inside MS = **one maneuver** (one discrete setting of the autopilot).
2) When we move inside MS
   a Moving **left** plans a **left turn** (decrease in heading)
   b. Moving **right** plans a **right turn**
   c. Moving **deeper in** plans a **throttle-in** (acceleration)
   d. Moving **out** plans a **throttle-back**
   e. Moving **up** plans a **climb**
   f. Moving **down** plans a **descent**
3) Colored maneuvers are unsafe (are predicted to → separation failure).
4) Color = available maneuver time (time left until that maneuver → separation failure).
   a. black represents no immediate conflict
   b. dark green represents 6 minutes to separation failure
   c. yellow represents 3 minutes to separation failure
   d. magenta represents 0 minutes to separation failure
   e. Our current autopilot settings = the exact 3D center of MS
5) No avoidance is needed unless MS center is colored (because MS center = current course). In our example, Figure 5 appears to be a close call requiring attention.
6) Maneuvers in conflict with more than one obstacle are colored for the conflict closest *in time*.
7) Maneuvers are translucent to allow us see through the entire MS.

In Figure 5, the entire greenish, translucent, glasslike structure is called a *conflict region* (CR). The CR represents all numerically contiguous unsafe maneuvers.

To restate the key concept, MS is a **maneuver-hypothesis tester.**

Potential human factors advantages of a MS-based CAS

Human behavior involves three fundamental components: perception, cognition, and action.

The essential perceptual advantage of MS may be that both unsafe and safe maneuvers “pop-out,” visually. “**Black = safe.**” “**Colored = unsafe.**” Such binary percepts can routinely be visually detected quite rapidly. For example, a familiar human face can be recognized as “familiar” (as opposed to not familiar) given stimulus presentation times of less than 50 ms (Carbon, 2003).

The essential cognitive advantage of MS may be that it transforms a difficult problem into an easy one. Just as binary percepts are easy to detect, so are many binary decisions easy to make (Smith & Ratcliff, 2004). Figure 2 looks difficult to solve. Nevertheless, in Figure 5, we can quickly see that a 1000-ft climb, or a 2° right turn, or a 5-kt speed increase would all put us back into black “safe” MS. A 1000-ft descent would be equally as safe (this is easy to see if we rotate the space and observe it from underneath). Moreover—assuming that the MS portrays “pop-out” solutions for all potential conflicts—this suggests that manually initiated solutions to even difficult traffic situations might rival the safety and efficiency of auto-CAS. Finally, if auto-CAS *were* in use, then MS should be capable of representing a system-planned maneuver on-hold, pending human approval.

An essential action advantage of MS is that it may solve the “triple-maneuver problem” of information representation, which currently plagues every maneuver-planning display so far conceived. Even maneuvers involving simultaneous heading+speed+altitude changes may be efficiently displayed and easily resolved in MS by simply moving a 3D planning cursor until it gets to a black “safe” region. Hitting an “Execute” button could then instantly reset the aircraft autopilot without need to dial in separate numbers. Since this preplanning can be done from the top level of the interface, there would be no informational graph structure to navigate, such as is typical of multifunction navigation displays.

Developing a MS-based CAS

For the current study, a collision avoidance system was developed to represent and use maneuver space. Figure 5 shows an annotated screenshot from the first such device.

In practice, this CAS requires position and velocity information for all aircraft of interest, a rotatable 3D display, a conflict probe to discriminate safe from unsafe maneuvers, a graphic method of drawing colored, translucent CRs, and a 3D cursor able to move within the MS, in order to plan and select a desired maneuver for execution.

Since available maneuver time adds a fourth dimension to the display, the entire CAS is technically 4D, hence its name 4CAS.

For initial proof of concept, a quasi-ballistic, deterministic, inertia-free conflict probe was used. Quasi-ballistic probes assume straight-line aircraft trajectories without knowledge of aircraft flight plan or pilot intent (climb and descent are accommodated, but leveling off and turns are not). Deterministic probes assume precise knowledge of aircraft position and velocity (as opposed to probabilistic probes, which assume a degree of error). Inertia-free probes assume instantaneous physical response of aircraft to control forces.

These simplifying assumptions may seem antithetical to proof-of-concept. However, many prototypes CASs begin the development cycle with simple conflict probes. The computational demands are less. Moreover, modern GPS allows extremely precise flight, most of which is straight-line/slow-change in en route airspace. Finally, initial tests of CASs typically focus on basic human factors/usability issues before moving on to complex engineering issues, since there is little sense investing time and money in any logically flawed system that operators inherently find hard to use.

Nonetheless, for the record, it is noted that full testing of an MS-based CAS will ultimately require a conflict probe capable of factoring in at least maneuver intent, type, and execution time. To many, the remaining issue of probabilism is moot, given modern cybernetic GPS guidance.

Anticipated economic benefits

This technology could afford three advantages. First, as noted, it ought to allow precise user-initiated navigation, even through “impossibly” crowded airspaces involving high traffic density, complex traffic geometry, and a wide range of aircraft speeds of the type we imagine in direct routing and/or pure, unrestricted free flight.

Second, it should be able to display auto-CAS maneuvers, allowing controller or pilot verification before a maneuver is initiated. This should facilitate user acceptance and allay any misplaced political fears over job losses due to automation. The human can be kept firmly in the loop, in tight supervision over the automation.

Third, the basic principle of MS is extensible to “cost space,” in which maneuvers are not only displayed in terms of safety, but also, literally, in terms of cost in time and/or money. While this is something to be explored later, it certainly is worth mentioning from the onset because the underlying concept is extensible to domains beyond aviation, making it appealing to the broader human factors community.

Proof-of-concept

All this is an interesting vision. However, the idea of a 4D nonveridical NAVAID is new to aviation. Basic usability issues have to be settled before further development is warranted. No new system can get by with just talking the talk; it also has to show it can walk the walk.

Version 1.0 of 4CAS was finished in the summer of 2004. As an "expert operator," the author could regularly solve difficult traffic situations and initiate successful, efficient avoidance maneuvers in 5-10 seconds. However, whether all users would find it that easy to use remained to be seen.

During the ensuing two years, 4CAS was set up as an experimental part-task simulator with data collection capability. Various revisions to the software were made. Version 1.1 was pretested with 18 general aviation (GA) pilots from Lund University, Sweden. Based on those results, further revisions were coded into V1.2. What follows is the first formal, published test of V1.2.

Experiment 1

Method

Participants. Eight male U.S. general aviation pilot volunteers participated with informed consent in Experiment 1 during the summer of 2006. Median age was 23 (range 22-34, mean 24.8, SD 3.9). Median civilian flight hours was 870 (range 250-1600, mean 891, SD 439). All participants held at least a private pilot's license; six held instrument ratings, six were double-certified as both Certified Flight Instructor (CFI) and Certified Flight Instructor-Instrument (CFII), and seven held Commercial ratings.

Apparatus. A part-task flight simulator was assembled, based on a Compaq Presario V2000 laptop computer with a 1.8 GHz AMD processor running Windows XP HE to a dual-head ATI Radeon Xpress 200M video card. The 4-Dimensional Collision Avoidance System (4CAS) and its companion cockpit display of traffic information (CDTI) were displayed on the laptop's native color monitor, while Microsoft Flight Simulator 2004 (FS2004) was displayed on an outboard Princeton LCD15 color flat-panel monitor. Both displays were run in 1024x768, 32-color mode, yielding on-screen widths x heights of 5 x 3.75 in (12.7 x 9.5 cm) for 4CAS, 4.8 x 5.3 in (12.2 x 13.5 cm) for the CDTI, and 12 x 9 in (30.5 x 22.9 cm) for FS2004.

FS2004 was set up to fly its Boeing 737-400 model, while its native Artificial Intelligence (AI) Traffic mode was used to create simulated en route air traffic. A shareware program, Traffic Tools V2.02, allowed partial control of this traffic by manipulating the departure and destination airports of the traffic as well as their takeoff times. This allowed rough setup of traffic position, speed, and heading. An interprocess communication program, FSUIPC V3.48, allowed 4CAS and FS2004 to talk back and forth. FS2004 allowed FSUIPC to read all own ship and AI Traffic variables critical to this experiment, namely latitude, longitude, heading, ground speed, and vertical speed.

Figure 6 shows a screenshot of 4CAS and the CDTI, 30 seconds into traffic scenario L045 (an approach from the left @ 45°).

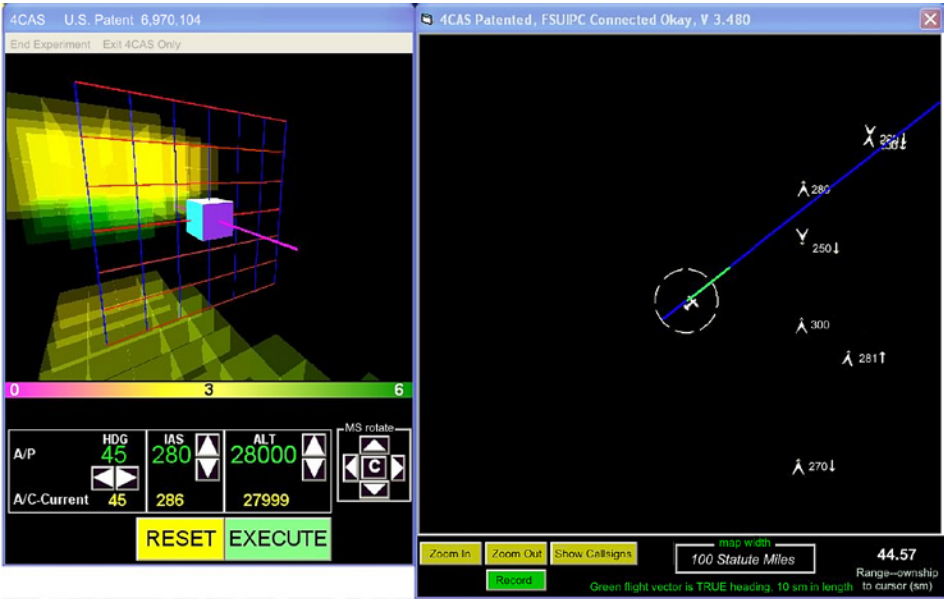


Figure 6. (Right) CDTI a view of scenario L045 from Experiment 1. (Left) 4CAS MS and CRs resulting from that traffic. The CRs show that there is no conflict right now. However, increased speed + a 1000' climb could *produce* a conflict in about 3 minutes. Turning left would produce a conflict in about 4.5 minutes, as would a 4,000' descent in about 4 minutes.

The CDTI showed the veridical, top-down, moving-map view of physical space. The own ship icon occupied display center, surrounded by a circular "J-ring" 5 nm in radius (9.3 km), representing the township's cylindrical protected zone, the safety buffer of inviolable airspace to be maintained at all times. Traffic was drawn as chevrons aimed in the direction of travel. An abbreviated text data tag showing aircraft call sign and traffic flight level (FL) was positioned close by. A "Hide call signs" button allowed removal of the call sign, leaving only the FL. Zoom buttons allowed selectable range views of 5, 10, 20, 40, 50, 100, and 200 sm (8, 16.1, 32.2, 64.4, 80.5, 161, 322 km). A green, 10-sm heading vector was drawn out from the own ship icon in the direction of travel, as was a blue path vector 10 minutes-equivalent in length (based on own ship ground speed). The blue path vector remained static relative to the absolute position of the virtual earth beneath, and ended with a red dot signaling the end of each experimental trial.

Normally, the CDTI updated and wrote data to file every 2 sec (2000 ± 5 ms). Any incursion of traffic within the own ship protected zone was treated as a *pilot deviation* (PD). A PD triggered a special burst timer, used to accurately capture point-of-closest-approach (PCA). This will be described in detail later.

4CAS normally stayed in plan mode. In plan mode, the MS could be rotated around either its vertical or horizontal axis, either by left-clicking and dragging with the optical mouse, or by the use of a control star (shown in Figure 6). In the center of the control star was a centering button labeled “C.” Clicking on “C” instantly redrew the MS and CRs, centering them to the standard frontal view.

The 3D cursor could be moved within the MS by the use of direction-control arrows, two for each axis. Within MS, left-right 3D cursor movement represented a planned heading change, up-down represented altitude change, and in-out represented speed change (“throttle-in, throttle-out”). If, during maneuver planning, the user wanted to cancel a planned maneuver, a “Reset” button quickly brought the 3D cursor back to MS center.

To resolve a conflict, no matter how intense or complex, the user had only to position the 3D cursor in a black region of MS and then hit the “Execute” button. This instructed FS2004 to immediately reset its autopilot and begin the maneuver.

While a maneuver was underway, the 3D cursor gradually “crept” back toward MS center. This was meant to emulate the idea of maneuver-in-progress. When the 3D cursor finally reached MS center, the CRs were redrawn to reflect a new MS centered on the new autopilot settings. The human factors implications of this method are discussed later. As mentioned, the 4CAS conflict probe was deterministic, quasi-ballistic, and inertia-free. The probe update rate was 2 sec.

Task. All scenarios began in mid-flight, at a nominal altitude of 28000 ft (FL 280) and indicated airspeed (IAS) of 280 kt. These emulated enroute free flight, in that aircraft were not restricted to normal odd-or-even flight levels by thousands (i.e., no East-West Rule).

The overall task was simply to stay on course—path initial altitude initial speed—deviating for traffic as necessary, then returning to course when clear of traffic. Upon reaching the “destination” after about 10 minutes, each scenario was ended manually.

Experimental Design. Experiment 1 was a repeated-measures design. Scenario presentation order was counterbalanced according to a standard 4x4 Latin square. Half the eight participants started in the CDTI-only condition, running the four scenarios in one of the four approach-angle sequences, followed by a short break, followed by the CDTI+4CAS condition, using the same scenarios and presentation order. The remaining participants ran similarly, but with the CDTI+4CAS first.

Participants were not told they would be repeating scenarios. Each 10-minute scenario effectively took about 14 minutes to run, including loading the input flight file, specifying the output data file, and letting each flight stabilize for 30 seconds before letting the pilot take over. This overall technique was found in previous research by Knecht and Hancock (1999) to minimize both scenario-specific learning and asymmetrical transfer (Poulton, 1982). Interposition of three other trials between the first and second exposure of duplicate scenarios—plus scenario setup time, plus the mid-experiment break—were all designed to let retroactive interference and time degrade memory of the first exposure.

Flight Scenarios. Four 10-minute scenarios were constructed. The number of traffic aircraft in each scenario varied dynamically, but typically maintained a light-to-moderate density of about 4-10 aircraft per 200x200 sm^2 maximum allowable zoom area on the CDTI. Figure 6 illustrates exactly what pilots saw 30 seconds into scenario L045 (an approach from the left @ 45º).

Each scenario represented crossing a bi-directional, vertical (north- or southbound) stream of traffic from one of four approach angles, 45, 135, 225, or 315 degrees (aeronautical coordinates, north=0, increasing clockwise). Xu and Rantanen (in press) suggest that these represent approach angles with relatively high conflict potential, although research on this is mixed (Nunes & Scholl, 2004).

FS2004 AI Traffic was programmed to fly straight and level. This is the simplest of all possibilities, a reasonable place to begin a series of experiments.

One scenario contained no prearranged traffic conflict (scenario L045). This was designed to test the pilots' false-alarm rate (avoidance maneuver in the absence of conflict). The remaining three scenarios contained one conflict each.

4CAS showed the maneuver space and conflict regions corresponding to the current traffic situation. Each CR cube showed heading-speed-altitude triplets predicted to intersect with traffic within the next 6 minutes. All CRs were translucent, allowing the user to see the 3D planning cursor even when obscured by other CRs. Each CR's color corresponded to the predicted available maneuver time (AMT) of traffic with the edge of the ship's own protected zone. CR colors were based on a lookup table of three anchor colors. Magenta anchored AMT=0 minutes, yellow anchored AMT=3, and green anchored AMT=6 minutes. Intermediate colors were linearly interpolated from the two nearest-anchor 0-255 integer RGB values. An AMT color reference bar was displayed on 4CAS, just under the MS.

In Figure 6, an upper-left set of CRs can be seen. This is actually two subsets, the higher, yellow subset corresponding to northbound traffic approaching the path from the right at FL300, and the lower, green subset corresponding to traffic, also northbound, approaching the path from the right at FL281, currently climbing at > 100 fpm but soon to level off at FL280. A lower-left, green set of CRs depicts northern traffic headed southbound at FL240. This cannot be seen on the CDTI at the zoom setting shown.

Dependent Measures. The first eight measures below, plus the latitudes and longitudes of the own ship and traffic, were recorded in real time for each pilot during each scenario, either every 2 sec, or immediately after an event occurred, whichever was appropriate. The final four measures came from Likert scales on a debrief form administered to each pilot after the experiment (see Appendix A).

•Maneuvers made:	Total number of maneuvers made during each scenario.
•Maneuver types made:	Number of maneuver types used per scenario (max = 3, the types being heading, speed, and/or altitude)

•Path length:	Total distance traveled (sm) during each scenario.
•3D max. deviation fr path:	3D normalized maximum deviation-from-flight-path during each scenario.
•Rmin:	Minimum 3D normalized range (Eq. 1) to closest traffic during each scenario.
•Maneuver onset time:	Elapsed time from start of scenario to first maneuver (seconds).
•PD duration:	Duration of each pilot deviation (seconds).
•PDs	Experiment-wide number of pilot deviations.
•Task ease:	Debrief question: During this experiment, how easy was it for you to avoid traffic?
•Time sufficiency	Debrief: Was there sufficient time to avoid traffic?
•Enjoyability Debrief:	How enjoyable was it to use the system?
•Training requirements:	Debrief: How many hours of training would you prefer before handling real traffic such as that experienced during the experiment?

During the tally of *Maneuvers made*, the definition of a single "maneuver" had to be operationalized to include sets of same-category maneuvers (*H, S,* or *A*), whenever two similar operations were executed within 8 seconds of one another on the FS2004 autopilot. This proved necessary because direct maneuver on the FS2004 autopilot often spanned more than the 2-sec data-write rate. In such cases, the pilot was obviously trying to execute a single maneuver; it was simply taking some time to click repeatedly on the autopilot. In addition, since maneuver was done directly on FS2004 mainly in the CDTI-only condition, if each recorded maneuver had been counted separately, this would have led to an artificial bias in favor of 4CAS showing fewer *Maneuvers made*. Therefore, the criterion of 8 sec was chosen to minimize such bias.

Path length was merely the sum of raw linear distances traveled from one data sample to the next. During deviation from path, the actual path deviated from the nominal path, increasing the total path length. However, note that altitude solutions tend to add less to total path length than heading solutions do. Therefore, *Path length* is somewhat biased in favor of altitude solutions.

3D Maximum deviation from path was defined as the length of a 3D vector drawn orthogonally from the actual path to the nominal path at the point of farthest deviation-from-path in normalized physical space. Normalization was defined in the standard mathematical sense, calculated by dividing lateral distances by 5 and vertical distances by 1000 ft to transform them into "standard units" in normalized 3D physical, "ATC space."

Rmin was based on logic similar to that just described. It was the minimum distance (point-of-closest-approach, PCA), in normalized physical space, from the own ship to the closest intruder during a single scenario

$$R_{\min} = \min\sqrt{\left(\frac{xy}{5}\right)^2 + \left(\frac{z}{1000}\right)^2}$$

where *xy* was lateral separation (nm) and *z* was horizontal separation (ft). *Rmin* was first described in Knecht and Hancock (1999). As we shall see later, *Rmin* may be slightly more complex than first envisioned, yet also somewhat more useful.

Maneuver onset time is a particularly important measure in conflict resolution, for two reasons. First, aircraft have substantial mass and limited power, yielding minimum times required to change speed or altitude. Second, heading solutions are easily achieved when far from an approaching obstacle, but become increasingly difficult during approach because the clearance angle required is $\alpha = \tan-1(y/x)$, x being the distance to the obstacle (a variable) and y the obstacle's half-width (fixed). Obviously, as $x \to 0$, α increases rapidly. Therefore, early maneuver is critical to both safety and efficiency.

Pilot deviations (PDs) were defined by the same standard as operational error in air traffic control, namely, violation of the own ship protected zone—approach of traffic to less than 5 nm lateral (xy) distance and less than 1,000 feet vertical (z) distance of the own ship while in enroute airspace. (FAA, 2006).

Burst timer. To facilitate accurate recording of aircraft separation during PDs, a burst timer was developed. This activated only when separation concurrently fell to less than 5 nm/1000 ft. During that time, as long as separation was decreasing, the burst timer sampled separation every 40 ms, but recorded to file only once every 2 sec. As soon as separation began increasing, the lowest value of separation captured thus far was immediately recorded to file. This allowed relatively sparse data collection under normal circumstances, keeping data files small, while still maintaining the ability to track minimum lateral aircraft separation to within accuracy of ±30 ft, even during a worst-case situation of traffic approaching head-on.

Training. Training was purposely kept brief, in order to explore inherent ease-of-use. It was important to explore how well minimally trained users could perform because, in actual free flight cockpit operations, users might only occasionally encounter conflicts, and might well be out of practice.

Pilots were first given a one-page instruction sheet describing their general task (i.e., to navigate safely through a bi-directional stream of traffic, generally staying on a blue path line at FL 280, deviating for traffic as necessary, then returning to course as soon as possible). They were then shown a one-page description of the CDTI and its operation. Those in the 4CAS+CDTI condition were then shown an additional one-page description of 4CAS and its operation.

Pilots were next allowed to practice on two training scenarios, being "walked through" the process of using the CDTI (and 4CAS, depending on to which half they were randomly assigned). They were offered the opportunity to practice as much as they wanted before starting data collection. Most elected to start after about 25-30 minutes of practice.

After completion of four data-collection scenarios, pilots were given a short break, after which they retrained for the second half of the experiment. Those who began with only the CDTI were given the instruction sheet for 4CAS. Everyone, no matter what their treatment order, was given the two practice scenarios again, before beginning data collection, and allowed to re-run those practice scenarios until ready to start data collection. Again, the typical re-training lasted about 25-30 minutes.

Results

Table 1 summarizes the relative performance of CDTI-alone trials versus CDTI+4CAS trials for the 8 participants x 4 trial-pairs = 64 total trials. Directionality of result is reported for all measures, significant or not, since the aggregate assessment of experiment-wide directionality across dependent measures constitutes information of interest (e.g., similar logic underlies the nonparametric binomial test)

Paired t-tests were used for continuous data, where a z-test of skew and kurtosis showed that to be appropriate (e.g., zskew = skew/SEskew). Standard errors came from Fisher (1925/1970). Otherwise, the nonparametric Wilcoxon test for paired-score ranks was typically used. For Pilot Deviations, a score of 1 was assigned to scenarios containing a PD, 0 otherwise. Scores were analyzed with the nonparametric McNemar test, since low expected cell values violated the assumptions of X2.

Angle-of-approach could not cleanly be tested for effect. Scenario L045 was confounded with also being the only non-conflictual scenario. In addition, angle was not independent of amount or positioning of traffic, since Traffic Tools did not allow sufficient precision. However, since scenario presentation order was counterbalanced, and all pilots experienced all scenarios, this was not expected to exert a significant experiment-wide effect.

Temporal effects. When maneuver onset time was regressed onto scenario presentation order, there were no apparent experimental effects merely due to the passage of time (e.g. practice, learning, or fatigue effects). This held true for CDTI-only trials (r2 = .001, ns), 4CAS-only trials (r2 = .091, ns), and all trials combined (r2 = .008, ns). Similar results were found for the other dependent measures.

Efficiency versus Safety versus Workload. Table 1 divides results into three categories of efficiency, safety, and workload. In theory, the safest aircraft are maximally separated, while the most efficient stick to their course. However, safety and efficiency are theoretically antithetical when traffic forces deviations from course. The best we can do is make necessary deviations as small as possible, given the safety standard (5 nm / 1000 ft).

High workload is theoretically antithetical to both safety and efficiency. When operators are stressed, we expect mistakes to be made.

Table 1

Experiment 1: Relative performance of CDTI-alone v. CDTI + 4CAS

	Dependent measure (1)	Units	Mean $\mu 1$ (CDTI-only)	Mean $\mu 2$ (CDTI + 4CAS)	p(skew normality) *distr1 , distr2*	p(kurtosis normality) *distr1 , distr2*	p (2-tail)	test	Significant?	Directionality favors 4CAS?
	Efficiency Measures									
1	Path length	sm	78.01	77.39	.009 , .169	.121 , .200	**.006**	Wilcoxon	Yes	Yes
2	3D max. deviation from course	SU (4)	1.57	1.37	.002 , .304	<.001 , .147	.280	Wilcoxon	No	Yes
3	*Rmin* (non-Pilot Deviations only)	SU	1.82	1.79	.362 , .364	.429 , .369	.693	t-test	No	Yes
4	False alarms (counts) (2)		N=7	N=6			1.0	McNemar	No	Yes
	Safety Measures									
5	*Rmin* (PDs only)	SU	1.21	1.34	.276 , .075	.190 , .374	.481	t-test	No	Yes
6	Maneuver onset time	sec	171.55	119.45	.301 , .014	.029 , .209	**.045**	Wilcoxon	Yes	Yes
7	Pilot Deviations, duration	sec	27.57	11.91	.003 , .023	.006 , .497	**.041**	Wilcoxon	Yes	Yes
8	Pilot Deviations (counts)		N=8	N=5			.508	McNemar	No	Yes
	Workload Measures									
9	Number of maneuvers made		6.09	4.63			.070	Wilcoxon	No	Yes
10	N. maneuver types made (3)		1.78	1.44			**.012**	Wilcoxon	Yes	Yes
11	Ease of avoiding traffic	1-6 Likert	3.9	4.9			.054	Wilcoxon	No	Yes
12	Had sufficient time to avoid traffic		4.5	5.4			.102	Wilcoxon	No	Yes
13	Enjoyability of use		3.6	5.3			**.033**	Wilcoxon	Yes	Yes
14	Amount of training required	hr	8.4	10.9			.458	Wilcoxon	No	No

(1) Measures 1-3, 5-7, 9, 10 compare trials within-pilots M 4, 8 are experiment-wide totals. M 11-13 are experiment-wide means on a Likert scale of 1-6, with higher numbers indicating 4CAS superiority. M 14 is estimated number of hours training needed to achieve competency.

(2) Non-conflict trials only (2 per pilot)

(3) Maximum of 3 per scenario (heading, speed, and/or altitude).

(4) SU = standard units (see Equation 1).

Efficiency

• Average path length	Significantly shorter with 4CAS present.
• Max. deviation from course	Not significant (*ns*), but directionality favors 4CAS.
• *Rmin*	*ns*, but directionality favors 4CAS.

When *Rmin* is used as an efficiency measure, only *error-free* scenarios are averaged (no PDs). In that case, "directionality" means "exceeding the safety standard, but by a smaller margin" (smaller *Rmin*), because additional, unnecessary separation would decrease fuel- and time-efficiency.

Safety

• *Rmin*	*ns*, but directionality favors 4CAS.
• Maneuver onset time	Significantly shorter with 4CAS present.
• Pilot deviations, duration	Significantly shorter, when PDs did occur.
• Pilot deviations, count	*ns*, but directionality favors 4CAS.

When *Rmin* is used as a safety measure, only *error* scenarios are averaged (those with PDs). In that case, "directionality" means, "having extra separation," (larger *Rmin*) because traffic is already violating airspace and needs more separation.

Workload

• N. maneuvers made/scenario	*ns*, but directionality favors 4CAS.
• N. maneuver types made/scen.	Significantly fewer with 4CAS present.
• Ease of avoiding traffic	*ns*, but directionality favors 4CAS.
• Time sufficiency	*ns*, but directionality favors 4CAS.
• Enjoyability of use	Significantly more enjoyable with 4CAS than without it.
• Amount of training required	*ns*, but slightly more training estimated to achieve proficiency with 4CAS+ CDTI.

It makes some sense that pilots would want more training with the 4CAS+ CDTI because there are two systems to learn rather than one.

Three debrief questions sampled the pilots' overall reaction to the CDTI and 4CAS (Appendix A). Based on a scale of 1 ("extremely unreasonable") to 6 ("extremely reasonable"), Q9 asked "are CDTIs a reasonable concept to continue researching (mean 5.1, SD, .64), while Q10 asked the same about 4CAS (mean 5.6, SD .52). The high means, low variability, and non-significant difference (p = .157, Wilcoxon 2-tailed) indicated that pilots valued both types of display highly and similarly. Finally, Q11 asked "If either system is not a good idea, which one, and why?" This drew just one response, "4CAS may draw attention away from the instrument scan"—a forthright answer, worthy of future study.

Discussion

Veridical displays represent realistic scenes. *State spaces* are *non*veridical displays used to visualize large amounts of data that would otherwise look unwieldy and confusing. *Maneuver space* (MS) is defined to mean a 4D nonveridical state space based on three coordinate axes of aircraft heading, speed, and altitude, plus a fourth dimension of available maneuver time. Maneuver space represents all conflictual and non-conflictual maneuvers achievable by one aircraft within a fixed look ahead time.

This work constitutes the first major successful formal test of a nonveridical, MS-based, 4D collision avoidance system called 4CAS.

Eight licensed general aviation pilots each flew eight simulated free flight scenarios, crossing a bi-directional traffic stream from various angles, using a standard cockpit display of traffic information alone or the same CDTI plus 4CAS.

Results suggested superiority of the CDTI+4CAS condition for one dependent measure of efficiency, two measures of safety, and two measures of workload. No practice or fatigue effects were seen. With 4CAS, flight path lengths were shorter and maneuvers were less complex and executed sooner, all with the same level of safety (if not better) than the CDTI alone.

All remaining dependent measures but one showed either trend or at least directionality in favor of 4CAS. A two-tailed binomial test of experiment-wide directionality (with 13/1 observed ratio versus 7/7 expected) yielded $p = .0018$. Such analysis is unconventional, however, and should be interpreted cautiously.

These results are modestly encouraging. They demonstrate that the idea of MS-based separation maintenance is operationalizable and intellectually indismissible.

However, the known limitations of this study do need to be made entirely plain. First, this was a “straw man” experiment. Just as a straw man is easy to knock over, an unadorned CDTI with no conflict alert or resolution capability is no match for a device having both. Consequently, this should be considered no more than the first in a series of increasingly challenging experiments to determine proof-of-concept. Second, naive operators were tested. While this is an essential part of testing, it says little about the asymptotic performance of highly trained users. Third, since the inventor also did the testing, there is always the possibility that pilots consciously or unconsciously set out to please the experimenter. Fourth, the system did not perform flawlessly. Improvements will be required. Finally, as is always true when testing a new paradigm, there are going to be unknown issues that arise unexpectedly. All we can do is to explore those in future experiments and report them as objectively as possible.

Lessons learned from Experiment 1

Experiment 1 revealed a number of ways to improve both testing and 4CAS itself, namely:

1. FS2004 AI traffic lacks the precision and controllability necessary to serve as a stand-alone traffic creation system for flight simulator experiments.
2. The Experiment 1 scenarios were not sufficiently challenging to elicit large differences in pilot deviations (PDs) between 4CAS and the CDTI.

3. The same scenarios contained a potentially trivial universal solution, namely to simply “dive beneath the traffic stream.”
4. The current version of 4CAS seemed to induce unnecessary maneuvers (false alarms) in near-conflict situations.
5. To maximize the accuracy of path length measures, program termination needs to be automated.
6. There may exist a fraction of pilots who misunderstand even the simple task presented here.
7. There may exist a fraction of pilots who fail to understand what 4CAS is, and how it works.

Lesson 1 revealed the inability of FS2004 AI Traffic to generate primary conflicts in precision experiments. First, it was possible to generate “traffic streams,” but not to control the exact behavior of any aircraft but the own ship. Second, traffic flew straight quite well but its vertical speed was uncontrollable and often unrealistically high (6000 fpm, in some instances). This led to disconcerting altitude “proposing” (oscillating overshoot/undershoot) and occasional conflict probe false alarms. This situation calls for a better method of traffic generation.

Lessons 2 and 3 involved traffic situations being too easy. Insufficient traffic density and overly simple traffic geometry were two reasons; another was the existence of a trivial maneuver solution—to simply descend until clear of all traffic. While only one pilot did this, it underscored the need for additional blocking aircraft.

Lesson 4 involved a high 4CAS false alarm rate (6/8 = 75%). For one thing, because the 3D cursor was opaque, it was often difficult to tell if conflict regions (CRs) actually overlapped MS center. Since this overlap was precisely the signal that a maneuver was necessary, this constituted a flaw calling for redesign. For another, the “creeping-cursor method” of representing real-time aircraft response turned out to have the somewhat paradoxical effect of looking like 4CAS was getting one *into* trouble, not out of it. This also called for redesign.

Lesson 5 involved manual shutdown of the program introducing a small error into the calculation of path length measures. This could be fixed by auto-shutoff at PCA-to-destination.

Lesson 6 involved a single pilot who misinterpreted the task, believing he was supposed to doggedly stick like glue to the blue path vector and altitude, and only deviate violently at the last possible minute. This left no choice but to test a substitute pilot and plan a rewrite of the experimental instructions to prevent such a misunderstanding from happening again.

Lesson 7 came from noting reversals of directionality in the recorded data (i.e. where performance on CDTI-only trials was better). These were infrequent, but naturally led one to question whether all participants truly understood 4CAS. This issue may merely turn on the quality of training. On the other hand, it may turn out that some fraction of the general population inherently has trouble with certain varieties of conceptual thought. This vital issue needs to be explored in future experiments.

One additional concern that has since been expressed about 4CAS involves the specific color scheme used to represent available maneuver time. For instance, if green is commonly used in other settings to represent no-conflict situations, why is it used here to represent conflict 6 minutes distant? The broad answer is threefold. First, “safe” MS is colorless for a reason. Black is the true “color” of empty space, plus non-zero R, G, B CR colors contrast well against black in a dim cockpit or room. Second, if “safe” MS *were* colored (e.g., green), then even a graphical rendering method involving thin mist or fog would needlessly obscure “distant” CRs. Finally, the “best” color scheme for AMT is a human factors issue to be determined by a combination of existing convention and empirical study. Current color standards are far from universal, and a body of evidence pertaining to multidimensional nonveridical aviation displays does not yet exist. Modesty reminds us that no invention is Athena springing fully formed from the forehead of Zeus. Prototypes evolve, as will this one.

Conclusions

Maneuver space data visualization is historically significant because this is the first time we are able to completely represent certain types of critical aircraft maneuver information simultaneously in a single display. A single MS-based display potentially contains everything we need to know to understand and resolve the large majority of air traffic conflicts, regardless of how complex their cause, or whatever that cause may be, be it traffic, terrain, special-use airspace, or weather. Any combination of simultaneous maneuvers can be depicted, up to and including heading+speed+altitude.

The only maneuver category MS cannot represent in a single display is segmented maneuver—multiple maneuvers executed serially. However, those pose a problem for all displays trying to display maneuver-combination solutions.

Maneuver space may prove particularly good at depicting maneuvers that efficiently satisfy safety requirements with minimum deviation-from-course. The MS is also amenable to transformation directly into cost space, and could graphically depict optimal maneuvers calculated by automatic collision avoidance systems. This would enable an air traffic controller or pilot to visually cross-check and approve maneuvers before initiation, maintaining ultimate human control while still enjoying the safety and economic benefits of auto-CAS.

In the end, the issue comes down to balancing simultaneous constraints of usability, safety, efficiency, system costs, and controller-pilot-passenger acceptance. Both veridical and nonveridical collision avoidance systems certainly have strengths. Hopefully, these are complementary. Maneuver space-based displays of air traffic information will never *replace* veridical displays—merely complement them.

Future experiments will attempt to remedy design and methodological deficiencies revealed here in Experiment 1. Experiment 2 is actually complete at the time of this writing, and addresses many of these issues.

Acknowledgments

This work was generously supported by a grant from the Wolf Aviation Fund. Dr. Kevin Williams wrote an ingenious CDTI that became the foundation for the one used here. Nicklas Dahlström and Jonas Dich of Lund University graciously pretested the first version of 4CAS with the help of 18 Swedish pilots. David McClurken, Chief Flight Instructor, University of Oklahoma School of Aviation, was instrumental in U.S. pilot recruitment. Deep thanks go to the American and Swedish pilots. Dr. Jing Xing supplied crucial insight on display color choice. Dr. Xidong Xu gave thoughtful comment on how best to convey many admittedly subtle concepts.

References

Bureau of Transportation Statistics. (2006). *Airline fuel cost and consumption (1977-2006).* Retrieved December 18, 2006 from www.bts.gov/xml/fuel/report/ src/index.xml.

Carbon, C.C. (2003). Face processing: Early processing in the recognition of faces. Unpublished doctoral dissertation, Freie Universität, Berlin.

Eby M.S., & Kelly, W.E. (1999). Free flight separation assurance using distributed algorithms. *Proceedings of the IEEE 1999 Aerospace Conference.* (pp. 429-444).

Federal Aviation Administration. (2006). *Order 7110.65R Air traffic control*. Washington, DC: Author.

Fisher, R.A. (1970) *Statistical methods for research workers*. Edinburgh: Tweeddale Court. (Original work published 1925).

Gibson, J.J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Hoekstra, J.M., Ruigrok, R.C.J., & van Gent, R.N.H.W. (2000, June). *Free flight in a crowded airspace?* Paper presented at the 3rd USA/Europe Air Traffic Management R&D Seminar, Napoli, Rome.

Johnson, W.W., Bilimoria, K.D., Thomas, L.C., Lee, H.Q., & Battiste, V. (2003, August). *Comparison of pilot and automation generated conflict resolutions*. Paper presented at the AIAA Guidance, Navigation, and Control Conference and Exhibit, Austin, TX.

Knecht, W.R., & Hancock, P.A. (1999). Separation maintenance in high-stress free flight using a time-to-contact-based cockpit display of traffic information. *Proceedings of the 4th Annual Meeting of the Human Factors and Ergonomics Society*. (pp. 16-20). Santa Monica, CA: Human Factors and Ergonomics Society.

Knecht, W.R., & Smith, K. (2001) The manoeuvre space. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics.* Aldershot: Ashgate.

Krozel, J. (2000). *Free flight research issues and literature search*. Los Gatos, CA: Seagull Technology.

Kuchar, J.K., & Yang, L.C. (2000). A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems. 1*(4), 179-89.

Nunes, A., & Scholl, B.J. (2004, March). *Interactions between convergence angle, traffic load, and altitude distribution in air traffic control.* Paper presented at the Third Annual Conference on Human Performance, Situation Awareness and Automation Technology, Daytona Beach, FL.

Operations Research and Analysis. (1998). *The impact of national airspace systems modernization on aircraft emissions*. (Technical Report DOT/FAA/SD-400-98/1). Washington, DC: Federal Aviation Administration.

Poulton, E.C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin, 91*(3), 673-690.

Radio Technical Commission for Aeronautics (RTCA). (January, 1995). *Report of the RTCA board of directors' select committee on free flight*. RTCA, Incorporated, Washington, DC.

Smith, P.L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences. 27*(3), 161-68.

Xu, X., & Rantanen, E.M. (in press). Effects of air traffic geometry on pilots' conflict detection with cockpit display of traffic information. *Human Factors.*

APPENDIX A

PARTICIPANT
NUMBER ____

1. How hard was it to avoid traffic conflicts using the CDTI alone?

Extremely hard	Quite hard	Somewhat hard	Somewhat easy	Quite easy	Extremely easy
1	2	3	4	5	6

2. How hard was it to avoid traffic conflicts using the CDTI + 4CAS?

Extremely hard	Quite hard	Somewhat hard	Somewhat easy	Quite easy	Extremely easy
1	2	3	4	5	6

3. Using the CDTI alone, did you feel you had enough time to choose a good maneuver and execute it?

Never	Rarely	Sometimes	Often	Always—barely	Always—easily
1	2	3	4	5	6

4. Using the CDTI + 4CAS, did you feel you had enough time to choose a good maneuver and execute it?

Never	Rarely	Sometimes	Often	Always—barely	Always—easily
1	2	3	4	5	6

5. Given the CDTI alone, how many hours of cockpit training would you want before handling real traffic like you saw?

1-2 2-4 4-8 8-16 16-32 32-64 more than 64

6. Given the CDTI + 4CAS, how many hours of cockpit training would you want before handling real traffic like you saw?

1-2 2-4 4-8 8-16 16-32 32-64 more than 64

7. How much fun was it to fly using the CDTI alone?

Never fun	Rarely fun	Occasionally	Frequently	Almost always	Always fun
1	2	3	4	5	6

8. How much fun was it to fly using the CDTI + 4CAS?

Never fun	Rarely fun	Occasionally	Frequently	Almost always	Always fun
1	2	3	4	5	6

9. In your opinion, are CDTIs a reasonable concept to continue researching?

Extremely unreasonable	Quite unreasonable	Somewhat unreasonable	Somewhat reasonable	Quite reasonable	Extremely reasonable
1	2	3	4	5	6

10. In your opinion, is 4CAS a reasonable concept to continue researching?

Extremely unreasonable	Quite unreasonable	Somewhat unreasonable	Somewhat reasonable	Quite reasonable	Extremely reasonable
1	2	3	4	5	6

11. If either system is not a good idea, which one, and why?

Private	Instrument	CFI I	CFII	Commercial	ATP
1 2	3	4	5	6	6

12. What ratings do you hold?

13. Age ___

14. Sex M F

15. How many total civilian flight hours do you have right now? (best guess) __________

16. How many total military flight hours do you have right now? (best guess) __________

17. What aircraft types have you flown? (the most frequently flown)

18. Describe how much other related experience you have, including computer games / flight simulators.

19. If you would like to participate in future experiments, please write your contact information on the back.

20. If you have any suggestions to improve this experiment, please write them on the back, also.

International Journal of Applied Aviation Studies, Volume 7, Number 1

Evaluation of Storm Forecast Displays for Air Traffic Control

Ulf Ahlstrom*
and
Ferne Friedman-Berg

Human Factors Team-Atlantic City
FAA Technical Center, Bldg. 28
Atlantic City International Airport, NJ 08405
USA

Ulf.Ahlstrom@faa.gov
Phone: 609-485-8642
Fax: 609-485-6218

Abstract

Adverse weather has a major impact on safety and operational efficiency during terminal air traffic operations. To mitigate these effects, researchers have suggested displaying weather information on controller displays. In the present study, we used simulation data to evaluate air traffic controllers' use of static and dynamic storm forecast displays that provided information about the direction of storm motion and future extrapolated positions (10-20 min). By analyzing eye-movement recordings, we made an assessment about controller fixation and scan path (a series of fixations and saccades) behavior. We found that the use of dynamic storm forecast displays significantly reduced controller scan path areas, scan path distances, and scan path durations compared to the static display, indicating greater efficiency of eye movements in the dynamic condition. In addition, the mean pupil diameter was significantly larger for controllers during the use of the static display compared to the dynamic display, indicating a higher visual and cognitive workload during the static condition. We discuss issues with displaying weather and traffic data on controller displays, as well as possible ways to improve storm forecast displays for tactical use.

Introduction

Adverse weather conditions have a major impact on air traffic operations, especially in the terminal domain. These conditions create safety hazards for pilots and constrain the usable airspace for Air Traffic Control (ATC). The result is reduced operational efficiency that often leads to delays and traffic diversions

---

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

Return to Table of Contents

(Sasse & Hauf, 2003). In an effort to mitigate these effects, there is an ongoing effort to improve the availability of weather information in ATC facilities. As part of this effort, current research has focused on designing information displays to present aviation and ATC related weather information (Ahlstrom, 2003).

In current ATC operations, air traffic management and supervisors use advanced weather information (e.g., storm motion predictions) primarily for planning purposes. They use this information strategically to foresee and manage operations, increase efficiency, and provide better service for the flying public during adverse weather conditions. The aim of this strategic thinking is to grasp the big weather picture, identify important long-term trends and patterns, and plan future actions.

In recent years, there has been an effort to evaluate the benefits and human factors issues associated with providing weather information directly to air traffic controllers. Although research has shown operational benefits from providing controllers with weather information (Ahlstrom, 2005a), interface designers still need to tailor this information to controller needs. This is because controllers must use weather information in an operational environment where the focus is on tactical, rather than strategic, thinking. Therefore, controllers will need weather displays that support the safe and efficient moment-to-moment management of air traffic within the airspace (Ahlstrom, 2005b). Such tactical weather displays might necessitate different display characteristics than the strategic weather displays currently used by meteorologists or by air traffic management.

Although strategic weather displays have been around for a long time, there is a need to adapt this information for use in tactical displays (Trafton & Hoffman, in press). When providing weather information for tactical and time-critical uses, interface designers must consider issues related to the temporal format (static versus dynamic) and display style (text versus graphics) of this information (Oron-Gilad, Meyer, & Gopher, 2001; Sanderson, Pipingas, Danieli, & Silberstein, 2003). Unfortunately, there is currently conflicting evidence regarding the ease-of-use for static and dynamic weather visualizations. Whereas weather forecasters can build dynamic mental models of weather patterns using static image information (Bogacz & Trafton, 2005), students of meteorology frequently have a difficult time in making effective use of static weather maps (Lowe, 2003). Additionally, when these students use dynamic weather visualizations, they typically extract information that is perceptually salient, but neglect the low-salience features even when those features convey important meteorological information.

A recent high-fidelity human-in-the-loop simulation evaluated operational benefits when controllers had access to both dynamic and static weather displays during severe weather avoidance (Ahlstrom, 2005a; Ahlstrom & Friedman-Berg, 2006). During all simulation runs, controllers had display access to six levels of precipitation information that showed the location and intensity of precipitation. During some runs, controllers also had access to static storm forecast (SSF) displays and dynamic storm forecast (DSF) displays (see Figure 1). The SSF display presented static weather information using solid or dotted lines and text, and showed the current storm cell positions and future extrapolated positions (10 - 20 min). The DSF display presented dynamic weather information by moving the solid display areas representing current precipitation cell positions to a future, extrapolated position (15 min).

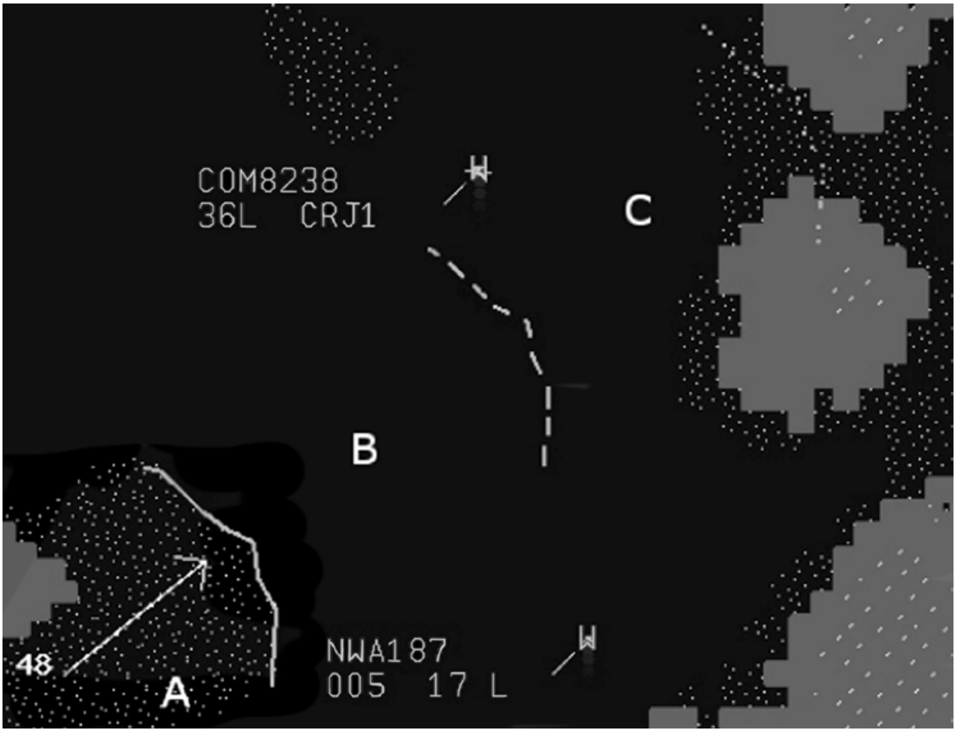


Figure 1. An illustration of precipitation levels and storm movement information used in the simulation. Precipitation levels 1-3 are shown in this illustration as light shaded areas, and level 4-6 are shown as dark shaded areas. For the display of the static storm forecast (SSF) display, the arrows show the direction of storm cell motion and the number indicates the storm cell speed in knots. Motion estimates were produced for all level 3 or greater cells and displayed on the area of heaviest precipitation within the cell. A solid line shows current storm cell position (leading edge), and extrapolated positions of the storm cell leading edge (10 and 20 minutes into the future) are shown by a broken and a dotted line, respectively. The dynamic storm forecast (DSF) display did not show any lines for current or extrapolated positions. Instead, this display showed a motion of the precipitation levels to a location 15 minutes into the future (i.e., to a location in between the 10 and 20 minute extrapolated positions for the SSF display). There were no other features for the DSF display. For our fixation analyses, we used the area covering the speed vector and numerals (A), the area between the storm cell leading edge and the 10 min extrapolated position (B), and finally, the area between the 10 and 20 min extrapolated positions (C).

Currently, SSF displays similar the one in Figure 1 are included in the Integrated Terminal Weather System (ITWS) that is used for strategic purposes by traffic management and supervisors in the terminal domain (Evans & Ducot, 1994). DSF displays are not available for current air traffic control operations, although similar dynamic displays are included in the Corridor Integrated Weather System (CIWS) currently under evaluation (Evans, Carusone, Wolfson, Crowe, Meyer, & Klingle-Wilson, 2002).

In the simulation conditions where controllers had access to storm forecast displays, researchers found an increase in traffic throughput (6% - 10%) and lower subjective workload ratings compared to conditions when controllers did not have access to storm forecast displays (Ahlstrom, 2005a; Ahlstrom & Friedman-Berg, 2006). Furthermore, controllers gave higher ratings of weather situation awareness, defined as the ability to predict and foresee how weather will affect future traffic, while using storm forecast displays. Although we found that the use of SSF and DSF displays increased-performance and decreased workload, the basic performance data and system measures did not capture any objective differences between the SSF and DSF displays. Therefore, there was a need to evaluate the use of these two displays in more detail.

The purpose of the present study was to evaluate the differences between SSF and DSF displays when used by air traffic controllers. For the simulation, we used an SSF display that already exists in the field; however, this display was not specifically designed for use with traffic data by air traffic controllers. Our DSF display, on the other hand, was created by a workgroup of TRACON controllers and research psychologists. The workgroup designed the DSF display to specifically improve upon the static storm motion display and create a tool that would be more useful for tactical use by controllers. This product does not currently exist in the field.

A promising technique for evaluating user interactions with displays is via the systematic analysis of eye-movement activity as users actively explore these displays (Brockmole & Henderson, 2005; Rayner, 1998). As a user interacts with a display, seeking out goal-relevant information, researchers can evaluate and compare their scan-path behaviors (i.e., ordered fixation/saccade patterns) for different display designs (Santella & DeCarlo, 2004). Even for very different information layouts, researchers can gain valuable insights into differences in information extraction by analyzing eye movements and scan paths (Rayner, Rotello, Stewart, Keir, & Duffy, 2001). Goldberg and Kotval (1999) showed that when information layout is good, viewers exhibit a close clustering in their fixation patterns, resulting in smaller scan path areas. However, when information layout is poor, fixations will be more diffuse, resulting in larger scan path areas.

We used eye-movement recordings to examine controllers' scan-path behavior during their use of SSF and DSF displays. We analyzed the location and duration of fixations and the distance of saccades to identify controller scan-paths as they interacted with each type of storm forecast display. Furthermore, we assessed the visual and cognitive workload during controller display interactions by recording their pupil diameter (Van Orden, Limbert, Makeig, & Jung, 2001).

By analyzing the controller scan pattern behavior for the SSF and DSF displays, we expect to learn how controllers used these displays during ATC operations. This information can help us improve the design and suggest necessary requirements for tactical forecast displays for controllers.

Method

Participants

Eleven non-supervisory, full-performance level Terminal Radar Approach Control (TRACON) controllers volunteered as participants in the simulation (*M* job experience =12 years [*SD* = 4.6]).

Simulation Design

Independent Variables. During the simulation, we used a 3 (Weather Display Location) x 2 (Weather Scenario) x 2 (Sector) repeated-measures design. Weather Display Location and Weather Scenario were within-subjects variables while Sector was a between-subjects variable. Because each controller only worked traffic in one of the two sectors, each controller participated in six scenarios. During the simulation, we counterbalanced the presentation order of the simulation conditions by means of a randomized block design.

The independent variable Weather Display Location created three different weather display conditions. In the first condition, *WIDS*, we presented the storm forecast displays (SSF and DSF) on an auxiliary Weather Information Display System (WIDS: Ahlstrom, Keen, & Mieskolainen, 2004) located on top of the controller workstation (see Figure 2). In the second condition, *workstation*, we presented both forecast displays directly on the controller workstation. In the third condition, the *control condition*, we did not present any forecast displays to controllers.

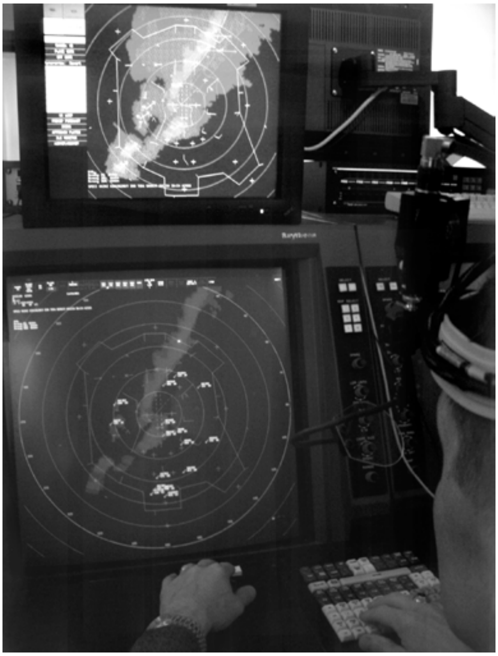


Figure 2. Simulation setup with the high-fidelity controller workstation (bottom) and the WIDS display (top). The researcher in the figure is wearing the Applied Sciences Laboratory Series 6000 Model 501 eye and head tracking system used during the simulation.

The purpose for manipulating Weather Display Location was twofold. First, it allowed us to evaluate the effects of providing controllers with weather displays on controller performance (e.g., traffic throughput). Second, it provided an opportunity to evaluate whether weather information in isolation (i.e., the WIDS condition) is more useful than weather information superimposed on traffic data (i.e., the workstation condition). Superimposing weather data on traffic data can be advantageous in that information can be viewed directly and relationally (Wickens & Carswell, 1995). However, it can also lead to display clutter and thereby make it more difficult for controllers to extract information (Yeh & Wickens, 2001). On the other hand, when weather information is presented separate from the traffic data (i.e., the WIDS condition), clutter is eliminated but the spatial separation-can reduce the ease by which controllers gain an overview of the traffic in relation to potential weather hazards (by forcing the controllers to mentally integrate the information on the two displays instead of viewing it directly).

The Weather Scenarios, the second independent variable, consisted of two sets of pre-recorded ITWS data from the Dallas-Forth Worth (DFW) TRACON. We used two storm scenarios with similar global motion (west to east) across the TRACON airspace. Each weather scenario was 50 minutes in length and contained a similar amount of line storm parts, rapid cell growth and decay, and a high degree of heavy (levels 3 and 4) and extreme (levels 5 and 6) precipitation cells.

The third independent variable, Sector, consisted of two adjacent TRACON sectors (West and East). We randomly assigned controllers to one of the two sectors and the controller worked the same sector during all simulation runs. We only recorded eye movement data from controllers working the West sector (referred to as West controllers).

Dependent Variables. During the simulation, we recorded dependent measures that corresponded to important ATC areas of severe weather avoidance, efficiency, safety, communications, weather situation awareness, workload, and the use of weather displays. We present these results in Ahlstrom (2005a) and Ahlstrom & Friedman-Berg (2006).

Simulation Apparatus and Procedures

We used a high-fidelity simulator that emulates the Standard Terminal Automation Replacement System (STARS) used in modern TRACON facilities, where air traffic controllers direct aircraft during the departure, descent, and approach phases of flight. We used a generic TRACON airspace (Guttman & Stein, 1997; Guttman, Stein, & Gromelski, 1995). The West controllers wore an oculometer consisting of an eye and head tracking system, Applied Sciences Laboratory Series 6000 Model 501 (Applied Science Laboratories, Inc., 2005), during all simulation runs.

Before every simulation run, we recalibrated the oculometer using a nine-point calibration grid. We used the Applied Sciences Laboratory software to enter the participant's point-of-gaze (POG) automatically for each of the nine calibration points. Using these known point locations, the software can determine the adjustments necessary to fit each POG to the exact location of the calibration points. After this calibration routine, we verified the calibration by having the participant refixate each point on the calibration grid.

Controllers received one day of training on the airspace and the use of weather information. At the start of the weather display training, controllers reviewed a training manual under the supervision of a TRACON subject-matter expert. After completing the training manual, controllers practiced using the weather displays on both the workstation and the WIDS display. Upon the completion of the weather display training, we found all controllers to be proficient in the use of the SSF and DSF displays. None of the training scenarios were used in the simulation.

To tailor the operating procedures for the simulation, we included a severe weather avoidance procedure that assigned responsibility for keeping aircraft away from precipitation levels 4, 5, and 6, to the controller. Controllers do not use this weather procedure in current field operations. It was included in the simulation to define controller responsibilities and to create an environment where the controllers had to use-weather displays.

Oculometer Recordings

Our head-mounted eye-tracker system integrates eye and head position data to measure a person's POG with respect to a fixed scene plane. During the simulation, we defined a calibration plane and four other bounded planes corresponding to the main situation display, the WIDS display, the flight strip bay, and the keyboard. For the present analysis, we only analyzed eye-movement recordings from the condition where controllers had access to storm forecast displays directly on their workstation (not the WIDS display). During the simulation, we recorded the eye's position in vertical and horizontal coordinates and the pupil diameter every 60th of a second (60 Hz). From these POG data, we calculated fixations, saccades, and the pupil diameter. Due to problems with track losses for one participant, we were only able to analyze the eye-movement data from five of the six participants. For a definition of how the software calculated these metrics, see Appendix A.

The Storm Forecast Displays

Both the SSF and DSF displays provided extrapolated positions for future storm cell locations. The SSF display provided default extrapolated positions for 10 and 20 minutes, whereas the DSF display used a default extrapolated position for 15 minutes. During the development of the DSF display, we used a workgroup of TRACON controllers and research psychologists to develop the display characteristics. While developing the display, the workgroup decided that the apparent motion sequence (produced by moving all the precipitation cells from their current to their extrapolated positions) was most optimal at a 15-minute extrapolation instead of a 10 or 20-minute extrapolation. Because the SSF display was based on an existing system, we chose not to change its existing extrapolated positions. This difference between the SSF and DSF display did not affect the ease by which controllers could judge extrapolated storm positions. Because all extrapolated positions are multiples of five, controllers could easily estimate the 5 or 15-minute position for the SSF by fixating a position between the middle of the leading edge and the 10-minute extrapolation, or between the 10 and 20-minute extrapolations. Controllers could perform a similar estimation procedure with the DSF display.

Upon activation by the controller, the system presented the SSF display until the controller deactivated the display. During development, we evaluated SSF displays that self-terminated after various durations. We found that the static display did not work well for the controllers when presented for short durations, and was too distracting if presented for longer durations. Therefore, we decided to let the controller activate and deactivate the display as needed. For our SSF display, all five features shown in Figure 1 were displayed simultaneously; there was no option for controllers to display individual elements selectively.

The activation of the DSF display resulted in an apparent motion sequence where the current storm cell positions moved to an extrapolated position 15 minutes into the future. In essence, the DSF display moved the precipitation cell areas instead of displaying the wind vector and solid/dotted lines for the current and extrapolated positions. Upon activation, the display moved all precipitation cells to the extrapolated position and displayed this position for 2 seconds before returning them to their original position. The controller used the trackball to control the DSF display duration. For each button press, the DSF display sequence continued for another two seconds, making it possible to view the sequence for as long as a controller deemed necessary. In contrast to the SSF display, we decided to use a fixed duration with automatic termination for the DSF display. The workgroup controllers reported that it was too distracting to have the display move back and forth continuously. Furthermore, we discovered during development and testing that controllers frequently used only one display sequence for the DSF display. During these single activations, controllers lost efficiency because they had to deactivate the display manually. If this would happen frequently during the simulation, it could have discouraged controllers from using the DSF display for a single activation. For simulation purposes, therefore, we decided to use a DSF display that terminated automatically after 2 seconds.

Analysis and Results

The data for the present analysis consists of weather display interactions and eye-movement recordings from the Ahlstrom and Friedman-Berg (2006) simulation. To optimize the analysis of eye-movement data, we only used recordings from the Weather Display Location condition where controllers had access to the weather displays directly on their workstation. In this condition, for a controller sitting at their position, the controller workstation covers a large viewing area (~ 38° x 38° of visual angle) and is at a straight viewing angle during ATC operations. Controllers also had access to a weather display during the WIDS condition. However, this auxiliary weather display covers a smaller viewing area (~ 30° x 24° of visual angle) and requires controllers to make larger head movements (i.e., the controller must look up at the display). Although this does not create a problem for controllers during WIDS operations, it is less optimal for recording eye movements. Therefore, in order to optimize the present analysis, we chose to use only the data from the West controllers in the condition where they used the weather displays on their workstation.

In the present paper, we analyzed the oculometer data only from the periods when controllers activated the SSF and DSF displays. First, we analyzed how frequently controllers activated the SSF and DSF displays and for how long these displays were shown once activated. Figure 3 shows the mean number of display activations (A) and the mean display duration (B) for the SSF and DSF displays, respectively. As shown in Figure 3, controllers activated the DSF display much more frequently than the SSF display ($t(4) = 2.24$, $p = .04$, Cohen's d[1] = 1.00, one-tailed). Furthermore, for each activation, controllers displayed the SSF for much longer durations than the DSF ($t(4) = 4.43$, $p = .005$, $d = 2.00$, one-tailed). On average, controllers only activated the DSF display twice in a row when used, producing average display durations of 4 seconds. When using the SSF display, however, controllers on average displayed the static information for more than 8 seconds, twice the average duration of the DSF display. Therefore, it seems that controllers were able to pick up information and focus their visual attention much more quickly when using the DSF display compared to the SSF display.

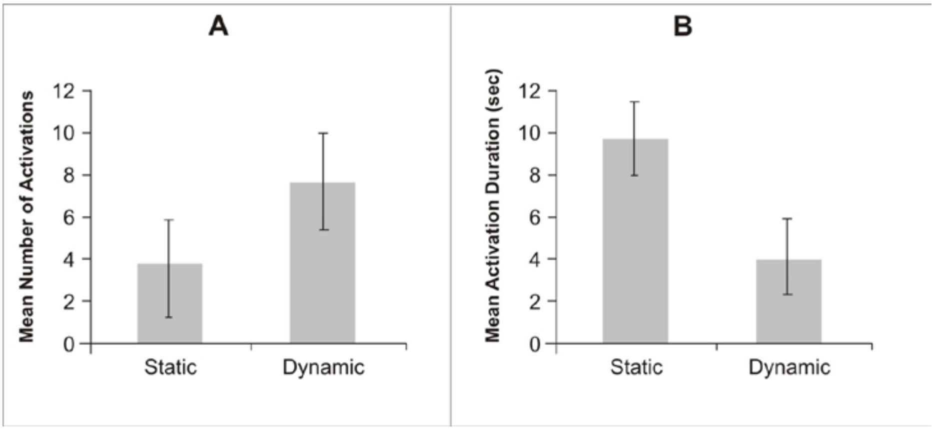


Figure 3. The mean number of activations (A) and the mean number of activation durations (B) for the static (SSF) and dynamic (DSF) display activations. The error bars are 95% within-subject confidence intervals (Cousineau, 2005).

Next, we attempted to determine whether any particular features on the SSF display captured controllers' visual interest. The DSF display did not have any distinct display features beyond the precipitation levels, and was therefore not included in this analysis. We wanted to know whether controllers fixated different elements of the SSF display in a fixed order, whether the controllers fixated certain display element features repeatedly, or whether fixation patterns were random. In addition, we were interested in whether controllers distributed their visual attention equally among the areas within the SSF display, or whether certain areas were of more visual interest than others were.

To identify all display elements that could have been the focus of attention during a single fixation, for each instance of the SSF display activation, we identified all display features within a two inch diameter of each fixation (i.e., the approximate foveal area), and ranked them in order from the nearest to the far-

---

[1] Cohen's *d* is an effect size measure that quantifies the magnitude of an effect (Rosnow & Rosenthal, 2003). For all calculations of *d* in the present study, we used the correction formula for an unbiased estimate of *d* proposed by Hedges and Olkin (1985).

thest. We then searched for fixation patterns in the data using the single feature nearest to the fixation (the most likely focus of attention). After identifying subject-specific patterns, we tried to identify fixation patterns that were similar across all six controllers. For these across-subject patterns, we calculated the average number of repetitions for each fixation pattern for each controller, averaging across only those controllers who exhibited that pattern.

Figure 4 shows the result of our fixation pattern analysis. We only found six 2-fixation patterns and one 3-fixation pattern in the data. The majority of these fixation patterns consisted of controllers looking at an aircraft, then fixating a SSF display feature, or fixating a SSF display feature, then fixating an aircraft.

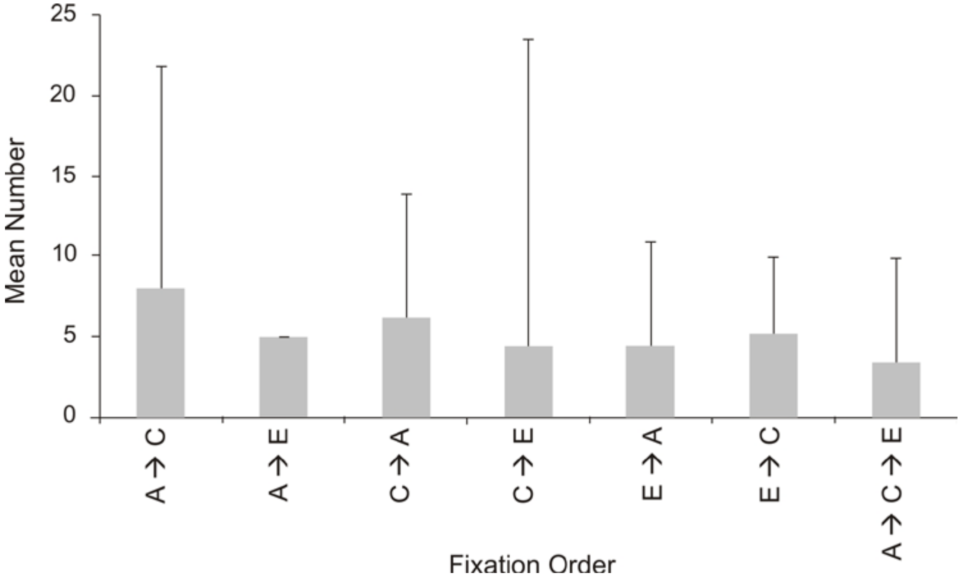
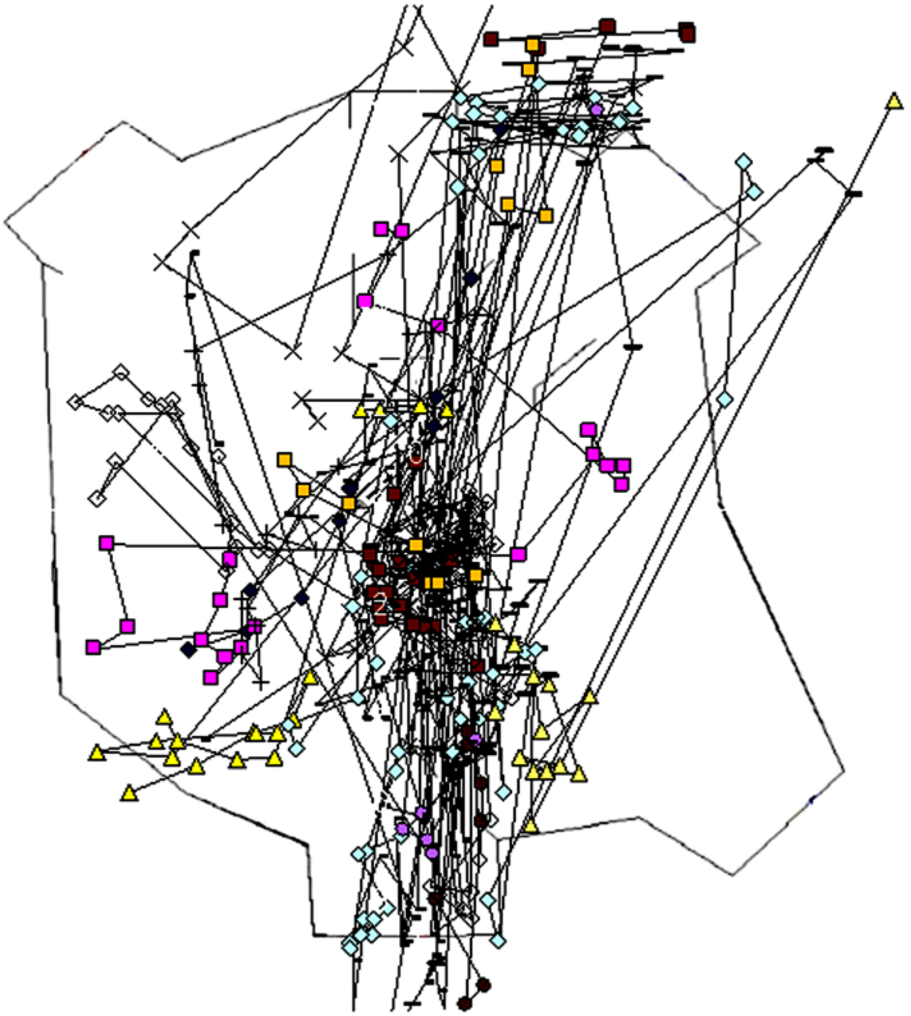


Figure 4. The mean number of fixation orders for the SSF display. The labels are A = Aircraft, C = Precipitation level, and E = Storm cell leading edge. The error bars are 95% within-subject confidence intervals.

For the 3-fixation pattern, controllers first fixated an aircraft, then a precipitation level, and finally a storm cell leading edge. These findings showed very little evidence of systematic fixation patterns to different SSF display features. We found no evidence of systematic higher-order (4 and 5) fixation patterns. The few systematic across-subject patterns that we found were 2-fixation patterns, and the vast majority of fixation patterns were unique to each controller.

We also analyzed the number and duration of fixations on specific SSF display features. We computed the mean number of fixations for the area covering the speed vector and numerals (A), the area between the storm cell leading edge and the 10 min extrapolated position (B), and finally, the area between the 10 and 20 min extrapolated positions (C) (see Figure 1 for reference). We found that across an entire simulation run, there were, on average, significantly more fixations in area B (*M* = 12.20, *SD* = 7.46) compared to areas A (*M* = 3.00, *SD* = 3.74) *t*(4) = 3.31, *p* = .03, *d* = 1.49, and C (*M* = 4.40, *SD* = 2.61) *t*(4) = 2.98, *p* = .04, *d* = 1.35, two-tailed tests. It appears that controllers focused their visual attention more frequently on the area between the leading edge and the 10 minute extrapolated position as compared to other SSF display areas. This result is consistent with subjective feedback from controllers, who stated that the SSF 10 minute extrapolated position was most useful for severe weather avoidance during the simulation. However, due to individual variations in fixation durations and a small sample size, we found no statistically significant differences in the mean fixation durations between these three areas (M_A = 284.11 msec, SD = 186.13; M_B = 442.79 msec, SD = 199.91; and M_C = 290.94 msec, SD = 61.18). Two-tailed tests for the difference in the mean fixation duration between area B and A showed $t(4) = 1.32$, $p = .26$, $d = .59$, and between area B and C $t(4) = 1.34$, $p = .25$, and $d = .60$. Nevertheless, effect sizes of $d = .59$ and $d = .60$ demonstrate non-trivial effects on fixation durations. Evidently, area B contained properties that caused controllers to focus on this area for longer durations than areas A and C.

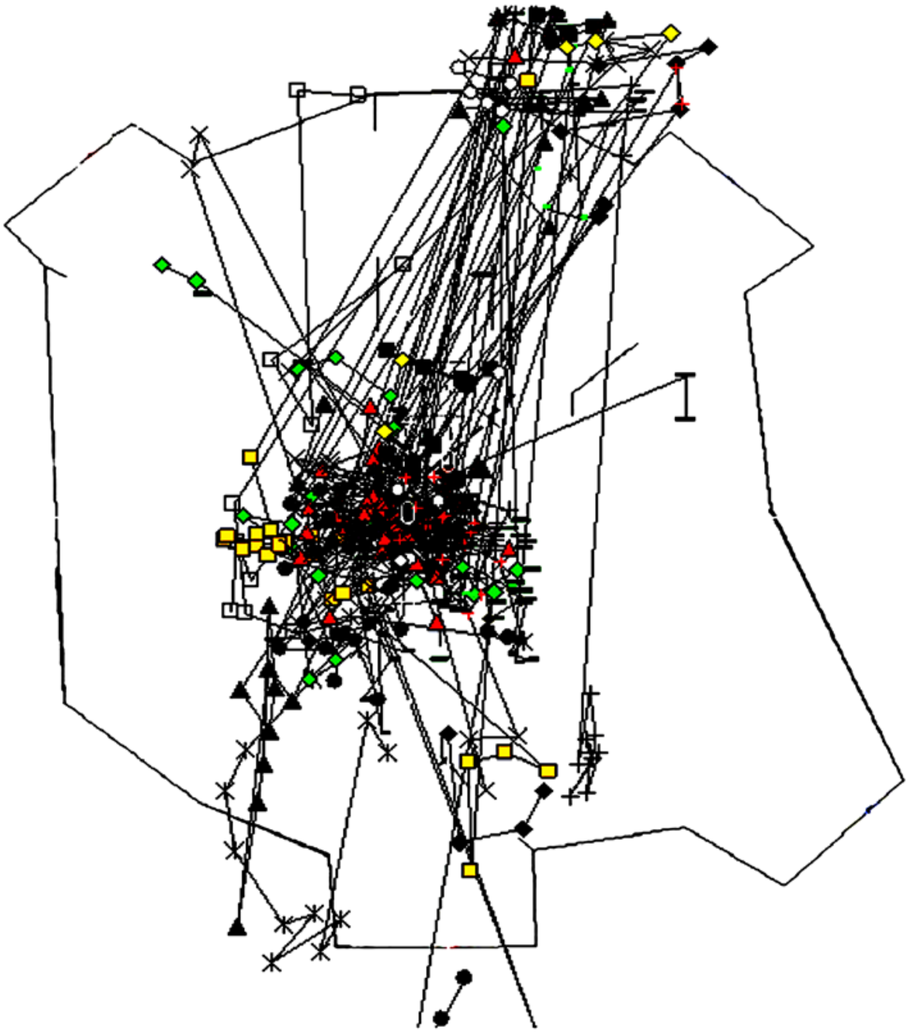
Next, we analyzed controller scan path behavior for SSF and DSF display activations by computing the scan path area, scan path distance, and scan path duration for each display activation. Each scan path consists of a series of fixations and saccades, and the size of the scan path area can be used to evaluate different display designs. Figure 5A and 5B show the actual scan paths for all controllers during the use of the SSF and DSF display respectively. In the figure, we have superimposed the scan paths on an outline of the TRACON airspace, with the West sector encompassing the left half of the airspace. The scan path patterns produced by the SSF and DSF displays show two distinct patterns. The SSF scan paths are fairly long and cover a large part of the TRACON airspace, with a small concentration in the area around the runways. The DSF scan paths are shorter, cover less area in the airspace, and have a lager concentration in the area around the runways. The individual controllers' scan paths also show the same distinct patterns. Figure 5C and 5D show the scan paths for a single controller, exhibiting the characteristic differences in scan paths produced by the SSF and DSF tools, respectively.



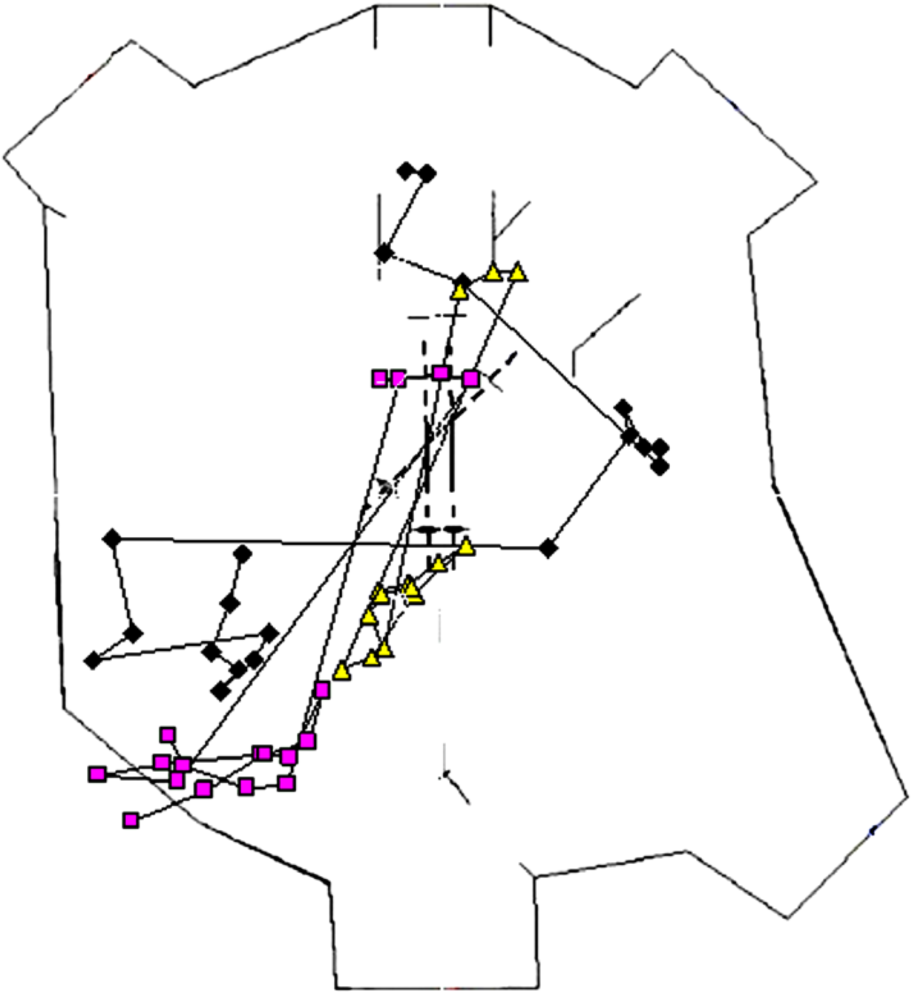
A

Figures 5 A-D. Illustrations of scan paths from five controllers during the use of SSF (A) and DSF (B) displays. In the figure, we have superimposed the scan paths on an outline of the TRACON airspace. The controllers only worked the West sector during the simulation (the left half of the airspace); the remaining airspace was outside the controllers' area of jurisdiction. The plots in C and D illustrate representative scan paths from a single controller during the use of the SSF and DSF displays respectively.

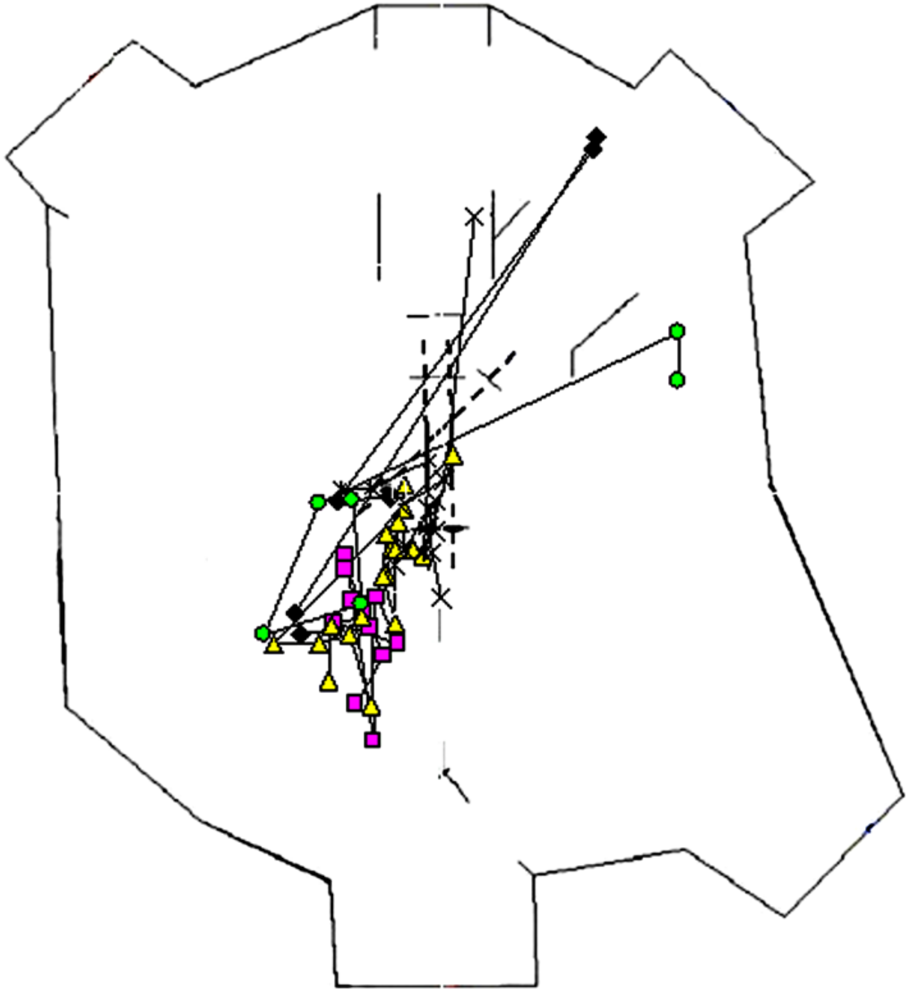
B



C



D



Goldberg and Kotval (1999) showed that when information layout is good, viewers exhibit a close clustering in their fixation patterns, resulting in smaller scan path areas. However, when information layout is poor, fixations will be more diffuse, resulting in larger scan path areas. Using the convex hull[1] function (Goldberg & Kotval, 1999) that defines the scan path area, we computed a convex hull for each display activation of the SSF and DSF displays. Figure 6 illustrates an example of the scan path area for a set of seventeen fixations (shown by dots). The line connecting these dots defines the convex hull for these fixations.

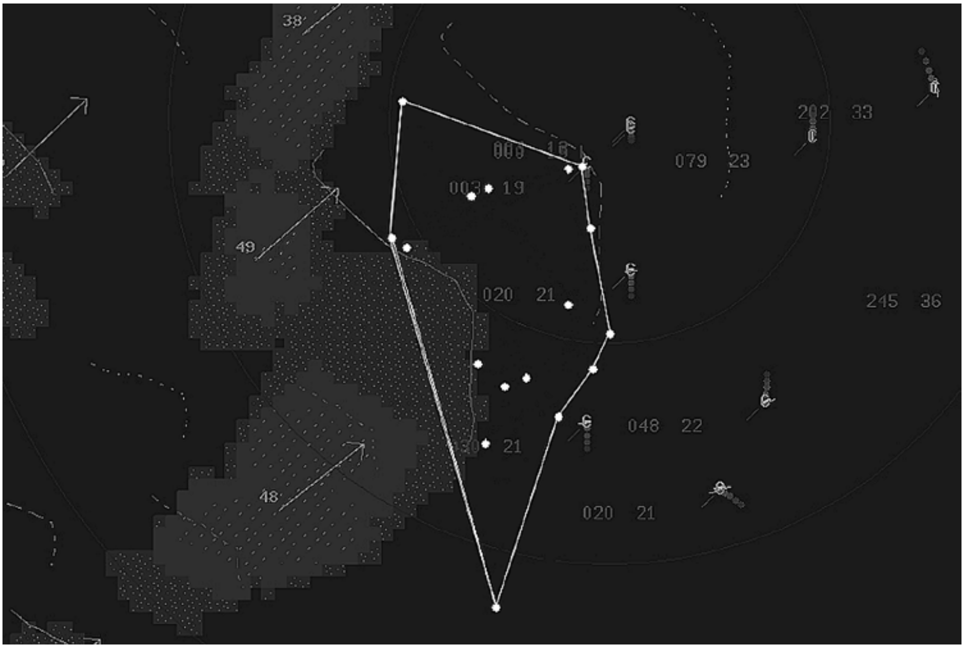


Figure 6. Illustrations of a controller scan path area during the use of the SSF display. The figure shows seventeen fixations (shown by dots) superimposed on the controller display. By computing the convex hull from the indices of the fixation points (shown by the line connecting the dots), we derived the scan path area.

Goldberg and Kotval (1999) also showed that the mean scan path distance and duration (the sum of all fixation and saccade distances and durations that make up a scan path) for well-designed displays are shorter than the mean scan path distance and duration for less well-designed displays, reflecting the degree of visual processing complexity. Previous research has also shown that different tasks that affect visual and cognitive workload can be measured by changes in ocular activity. Pupil diameter in particular has been shown to increase with increasing visual and cognitive demands. Van Orden et al. (2001) found an increase in pupil diameter with increasing display target density during a target identification task. Therefore, we performed an analysis of the pupil diameter for each controller as an indicator of relative visual and cognitive workload during his or her use of the SSF and DSF displays.

Figure 7 shows the mean scan path area (A), the mean scan path distance (B), the mean scan path duration (C), and the mean pupil diameter (D) for the SSF

---

[1] The convex hull of a set of points is the smallest convex set that includes all the points. It can be described as the set of convex combinations of the set of points of the form $\boxed{\sum_{j=1}^{n} t_j x_j}$, where n is an arbitrary natural number, the numbers t_j are non-negative and sum to 1, and the points x_j are in X.

(static) and DSF (dynamic) conditions. As illustrated in Figure 7A, the mean scan path area for the static display was significantly larger than the scan path area for the dynamic display ($t(4) = 4.07$, $p = .01$, $d = 1.84$, two-tailed). Activation of the static display appeared to result in a more widely distributed scan path. As stated previously, this is a phenomenon usually interpreted as being an artifact of poorly designed interfaces (Goldberg & Kotval, 1999). Figure 7B shows that the mean scan path distance after the activation of the static display was significantly longer than the mean scan path distance after the activation of the dynamic display ($t(4) = 4.78$, $p = .008$, $d = 2.15$, two-tailed). This indicates that controllers had more optimal scan paths when using the dynamic display than when using the static display. Figure 7C shows that the mean scan path duration after the activation of the static display was significantly longer than the mean duration after the activation of the dynamic display ($t(4) = 4.43$, $p = .01$, $d = 2.00$, two-tailed), indicating that there was greater difficulty in extracting goal-relevant information from the static display than the dynamic display. It should be noted that the mean scan path durations in Figure 7C are identical to the mean activation durations in Figure 3B. Because the scan path duration is a sum of all fixation and saccade durations, and given the fact that controllers were fixating the workstation when they activated the SSF display, these values are identical in our data. Finally, Figure 7D shows that the mean pupil diameter was significantly larger after the activation of the static display than after the activation of the dynamic display ($t(4) = 5.54$, $p = .005$, $d = 2.50$, two-tailed). Evidently, using the static display increased visual and cognitive demands more than using the dynamic display.

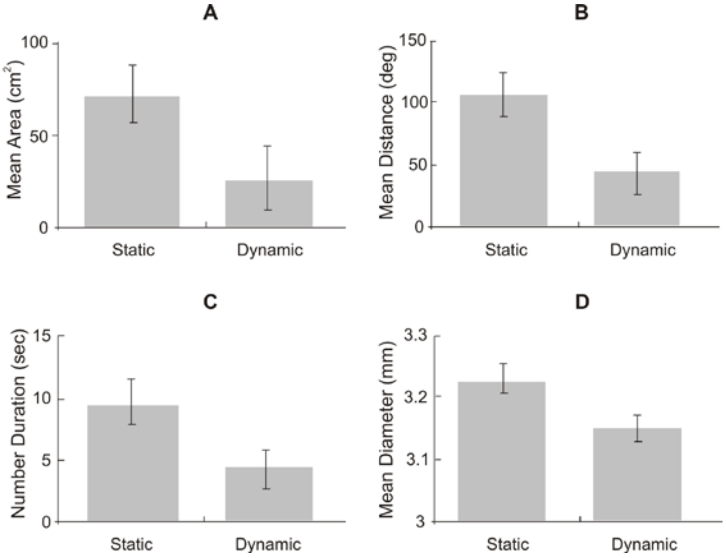


Figure 7. (A) Mean scan path area in cm^2, (B) mean scan path distance in degrees of visual angle, (C) mean scan path duration in seconds, and (C) mean pupil diameter in millimeters for the static (SSF) and dynamic (DSF) display conditions. The error bars are 95% within-subject confidence intervals.

Discussion

The present study suggests that although static weather displays seem more efficient for strategic use by meteorologists (Bogacz & Trafton, 2005) and air traffic management, static storm forecast displays are less than optimal for use in a tactical domain like air traffic control. Although air traffic management and controllers use weather information to avoid hazardous weather areas, there is an important difference in their decision-making strategies. Air traffic managers and supervisors make strategic decisions when planning for operations, while air traffic controllers make near-term tactical planning decisions. Strategic decision-making is less time critical than tactical decision-making. There is also a higher degree of multitasking involved in tactical (i.e., moment-to-moment) decision-making compared to strategic decision making. Therefore, controllers need information displays that promote effortless information pick-up to avoid increased workload. Dynamic weather displays might be more suitable for controllers because they eliminate the need to move storm cells mentally while performing ATC tasks.

Bogacz and Trafton (2005) suggest another reason why dynamic weather displays might be more suited to air traffic controllers than static displays. They claim that meteorologists, because of their expertise, knowledge, and ability to use critical weather features, are capable of mentally animating static displays, thereby alleviating the need to display actual animations. Controllers, on the other hand, react to time-critical situations and must be able to extract goal-relevant information about weather and traffic directly from their display. For air traffic controllers, it is more efficient to have access to display information that directly specifies control opportunities (i.e., affordances) without any need for mental elaborations (Ahlstrom, 2003).

Our scan path analysis also supports our conclusion that dynamic forecast displays are more appropriate for controllers than static forecast displays. For example, as a controller views an aircraft that is approaching the runway during a storm, the controller has an immediate need for information about storm movements to ascertain that no future conflict between the airplane and storm cells will occur. If the controller activates a static storm forecast display, the system will present information about current and extrapolated storm cell positions. In this instance, optimally designed weather displays would facilitate goal-directed fixation behavior, resulting in a close clustering of fixations and saccades within a relatively small scan path area that is directly relevant for the task.

We found quite the opposite result for the SSF display in the present analysis. Rather than facilitating and enhancing controllers' fixation behavior, the current design of the static display produced less efficient scan paths. Because the onset of SSF features was abrupt and spread across the entire display rather than localized to goal-relevant areas, it appeared to capture controllers' visual attention. This is known as the attention capture phenomenon (von Mühlenen, Rempel, & Enns, 2005). It implies that some events (e.g., abrupt onset or feature change) penetrate attention even though they might be completely irrelevant for the task. Von Mühlenen et al. (2005) suggested that attentional capture is the result of a conflict between an observer's immediate task goals and the visual system's priority for novelty detection. Use of the DSF display resulted in significantly less attentional capture than the SSF display, possibly because of the common image motion (*common fate*) in the dynamic display compared to the diffuse nature of the elements in the static display. If the visual system has a tendency to view the features of the static display as separate elements, then upon its activation, there will be more novel elements to detect than upon the activation of the common fate movement (Ahlstrom & Borjesson, 1996) of the dynamic display, resulting in greater attentional capture. This supports our view that dynamic forecast displays might be more appropriate for visuospatial tasks like tactical air traffic control.

Although the abrupt onset of display features can lead to attentional capture, it is possible that we could design SSF displays that produce less attentional capture. For example, we could display SSF information on smaller but more goal-relevant areas of the situation display. In our simulation, controllers used the SSF and DSF displays for specific purposes like severe weather avoidance during the timing of arrivals or for runway selection. If we restricted the display of SSF information to task-relevant display areas, effectively reducing the number of novel elements on the display, we could limit the abrupt onset of global display information and could potentially reduce the negative effect of attentional capture. Furthermore, by only displaying elements that define the area between the storm cell leading edge and the 10 minute extrapolated position, which is the area controllers fixated most and reported as being the most useful, we could further reduce the number of novel elements, which might enhance controller fixation behavior even further.

Based on the present analysis, it appears that dynamic displays are more effective than static displays for increasing controller scan-path efficiency. From an operational perspective, any storm forecast display that requires, on average, nine seconds to extract information can be detrimental to controller performance. This is time not fully spent concentrating on sector traffic. However, if we modify the static displays as suggested above it may minimize this difference. Although these display suggestions are by no means exhaustive, we do believe that by using them in the design of future weather displays, we have the potential to create storm forecast displays that increase controller efficiency while reducing visual and cognitive workload during tactical ATC.

References

Ahlstrom, U. (2003). Current trends in the display of aviation weather. *Journal of Air Traffic Control, 45*(3), 14-21.

Ahlstrom, U. (2005a). Advanced weather displays for TRACON controllers. *Journal of Air Traffic Control, 47*(2), 29-36.

Ahlstrom, U. (2005b). Work domain analysis for air traffic controller weather displays. *Journal of Safety Research, 36*(2), 159-169.

Ahlstrom, U., & Borjesson, E. (1996). Segregation of motion structure from random visual noise. *Perception, 25,* 279-291.

Ahlstrom, U., & Friedman-Berg, F. (2006). *TRACON controller weather information needs: III. Human-in-the-loop simulation* (DOT/FAA/TC-06/10). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ahlstrom, U., Keen, J., & Mieskolainen, A. J. (2004). Weather Information Display System (WIDS). *Journal of Air Traffic Control, 46*(3), 7-14.

Applied Science Laboratories, Inc. (2005). *Series 6000 Model 501 Eye Tracking System with Head Mounted Optics* [Computer software]. Applied Science Laboratories, Inc., Bedford.

Bogacz, S., & Trafton, G. (2005). Understanding dynamic and static displays: using images to reason dynamically. *Cognitive Systems Research, 6*(4), 312-319.

Brockmole, J. R., & Henderson, J. M. (2005). Prioritization of new objects in real-world scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 857-868.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology, 1*(1), 42-45.

Evans, J. E., Carusone, K., Wolfson, M., Crowe, B., Meyer, D., & Klingle-Wilson, D. (2002). The Corridor Integrated Weather System (CIWS). In *Proceedings of the 10th Conference on Aviation Range and Aerospace Meteorology*. American Meteorological Society, Portland, Oregon. Retrieved January 26, 2006, from http://www.ll.mit.edu/AviationWeather/Evans-WW-10473.pdf.

Evans, J. E., & Ducot, E. R. (1994). The Integrated Terminal Weather System (ITWS). *The Lincoln Laboratory Journal, 7*(2), 449-474.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics, 24*(6), 631-645.

Guttman, J. A., & Stein, E. S. (1997). *En route generic airspace evaluation* (DOT/FAA/CT-TN97/7). Atlantic City International Airport: Federal Aviation Administration William J. Hughes Technical Center.

Guttman, J. A., Stein, E. S., & Gromelski, S. (1995). *The influence of generic airspace on air traffic controller performance* (DOT/FAA/CT-TN95/38). Atlantic City International Airport: Federal Aviation Administration William J. Hughes Technical Center.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction, 13*(2), 157-176.

Oron-Gilad, T., Meyer, J., & Gopher, D. (2001). Monitoring dynamic processes with alphanumeric and graphic displays. *Theoretical Issues in Ergonomics Science, 2*(4), 368-389.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*(3), 219-226.

Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, *57*(3), 221-237.

Sanderson, P., Pipingas, A., Danieli, F., & Silberstein, R. (2003). Process monitoring and configural display design: A neuroimaging study. *Theoretical Issues in Ergonomics Science, 4*(1-2), 151-174.

Santella, A., & DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the Eye tracking research and applications symposium on eye tracking research and applications*. ACM Press, New York, pp. 27-34.

Sasse, M., & Hauf, T. (2003). A study of thunderstorm-induced delays at Frankfurt Airport, Germany. *Meteorological Applications, 10*, 21-30.

Trafton, J. G., & Hoffman, R. R. (in press). Computer-aided visualization in meteorology. *IEEE Intelligent Systems*.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43*(1), 111-121.

von Mühlenen, A., Rempel, M. I., & Enns, J. T. (2005). Unique temporal change is the key to attentional capture. *Psychological Science, 16*(12), 979-986.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundations and its relevance to display design. *Human Factor*s, *37*(3), 473-494.

Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding, and decluttering techniques. *Human Factors*, *43*(4), 543-562.

Appendix A:

Eye tracking metrics

The Research, Development, and Human Factors Laboratory uses the Applied Science Laboratories Model 6000 oculometer consisting of an eye and head tracking system (Applied Science Laboratories, Inc., 2005). The system measures both eye and head movement at 60 Hz to record points-of-gaze (POG) in x, y, and z coordinates relative to the defined scene planes (e.g. the radar display, the keyboard). Visual angle is determined by the distance between POGs on the scene plane (x, y) and the distance between the observer's head and the scene plane (z). The eye tracker analysis software uses the visual-angle-based algorithm to identify fixations and saccades, to define their durations, and to calculate other metrics like blink frequency and duration and pupil diameter.

Fixation

A fixation is a sequence of at least six oculometer samples with an intersample distance of less than 1 degree of visual angle. At 1-meter distance, this corresponds to a circle with an 8.73 mm radius. The distance between two samples is the norm of the vectorial difference of the sample coordinates. If two fixations are not separated by either a blink or a saccade (see definitions below), these fixations should be combined within one fixation. In summary:

Fixation if:

$$D = \sqrt{((x_i-x_{i+1})^2 + (y_i-y_{i+1})^2)} > 8.73 \text{ mm}$$

with D the distance between to subsequent samples x and y the horizontal and vertical point of gaze coordinates in mm respectively

and:

$n > 6$ with n the number of samples in a sequence

and

separated by a blink or a saccade

Related to a fixation the following variables need to be calculated: Fixation Duration and Fixation Area. Fixation Area is an approximation of the area covered by the POG due to eye movements within a fixation.

Fixation Duration:

$$\mathrm{FIXDUR} = t_{sample} * \Sigma \mathrm{samples}$$

with t_{sample} where the duration of a sample ($^{1}/_{60}$ second) and Σsample is the total number of samples within a fixation

Fixation Area:

$$\mathrm{FIXAREA} = (\max(x_{fix})-\min(x_{fix}))*(\max(y_{fix})-\min(y_{fix}))$$

with x_{fix} and y_{fix} the sequences of horizontal and vertical POG coordinates within a fixation respectively

Blink

A blink is the complete or partial closure of the eye. The oculometer will suggest that the velocity at the start and end of a blink was greater than 700 degrees per second, which corresponds with 6.108 $^{m}/_{s}$. This is physically impossible, but it does give us a way to determine the start and end of a blink. A blink starts after the last sample of the previous fixation and stops before the first sample of the next fixation. In summary:

Blink if:

$$\mathrm{VEL} = \sqrt{((x_i-x_{i+1})^2 + (y_i-y_{i+1})^2)} / t_{sample} \quad > 6.108\ ^{m}/_{s}$$

with VEL being a crude estimate of the tangential velocity and x and y the horizontal and vertical point of gaze coordinates in mm respectively. The index denotes the current sample i and next sample i+1 respectively

and:

$$n > 12$$

with n the number of samples in a sequence

Related to a blink the following variables need to be calculated: Fixation Duration and Blink Distance. Blink Distance is the distance covered by the POG due to eye movements during a blink.

Blink Duration:

$$\mathrm{BLNKDUR} = t_{sample} * \Sigma \mathrm{samples}$$

with t_{sample} where the duration of a sample ($^{1}/_{60}$ second) and Σsample is the total number of samples within a blink

Blink Distance:

$$\mathrm{BLNKDST} = (x_n-x_p)*(y_n-y_p)$$

with x and y the horizontal and vertical point of gaze coordinates in mm respectively. The index denotes the last sample of the previous fixation p and first sample of the next fixation respectively

Saccade

A saccade is the ballistic movement of the eye from one fixation to the next. A saccade is characterized by fast eye movements of up to 700 degrees per second. The cut-off for a saccade is a difference in distance between two subsequent saccades that is greater or equal to 8.73 mm, lasts at least 3 samples (or a velocity of 0.524 $^{m}/_{s}$), and the velocity is less or equal to 700 degrees per second (6.108 $^{m}/_{s}$). The saccade will start at the end of the last sample of the previous fixation and will end at the beginning of the first sample of the next fixation. In summary:

and: $0.524 > VEL > 6.108\ ^{m}/_{s}$

$n > 2$

Related to saccades a number of variables need to be calculated: Saccade Duration, Saccade Distance, and Saccade Velocity. The saccade distance is the angular distance traveled during a saccade in degrees. The saccade velocity is the average velocity within a saccade in degrees per second.

Saccade Duration:

$$SACDUR = t_{sample} * \Sigma samples$$

with t_{sample} where the duration of a sample ($^{1}/_{60}$ second) and Σsample is the total number of samples within a saccade

Saccade Distance:

$$SACDST = (x_n-x_p)*(y_n-y_p)$$

with x and y the horizontal and vertical point of gaze coordinates in mm respectively. The index denotes the last sample of the previous fixation p and first sample of the next fixation n respectively

Saccade Velocity:

$$SACVEL = \Sigma\ (\sqrt{}((x_i-x_{i+1})^2 + (y_i-y_{i+1})^2))\ /\ t_{sample} * n_{saccade}$$

with t_{sample} where the duration of a sample ($^{1}/_{60}$ second) and $n_{saccade}$ is the number of samples within the saccade

Pupil diameter

Pupil diameter is calculated by knowing the distance of the participants' eye to the oculometer camera. The conversion of the relative pupil diameter output to millimetres is performed in three steps. First, an ASL Model Eye (Applied Science Laboratories, Inc., 2005) is placed at the same eye to camera distance as present during the simulation conditions. Second, a proper discrimination is obtained on the model pupil and the corresponding pupil diameter value is noted. Third, a scaling factor is computed by dividing the recorded value by 3.96 (the diameter of the Model Eye pupil) and then multiply this scaling factor by the recorded pupil diameter value (value in millimetres = scale factor * recorded value).

International Journal of Applied Aviation Studies, Volume 7, Number 1

A Multidimensional Scaling and Participatory Design Approach to Classify Open-Ended Aircraft Maintenance Data

Kunal Kapoor
and
Joel S. Greenstein

147, Freeman Hall,
Human Computer Systems Laboratory,
Clemson University,
Clemson, SC 29634
Phone: 864-656-4717
kunalkapoor1975@gmail.com

Abstract

The quality assurance data to be analyzed by the web-based surveillance and auditing tool (WebSAT) is both qualitative and quantitative. The forced responses to checklist questions provide a definitive outcome identifying the effectiveness of the four quality assurance work functions. On the other hand, open-ended responses, the second type of response for capturing maintenance errors, are qualitative in nature since they reflect what the auditors and quality assurance representatives observe during their interactions at vendor locations. This research proposes to apply the statistical technique of multidimensional scaling (MDS) and the User Centered Design (UCD) method of Participatory Design (PD) to categorize open-ended responses into suitable performance metrics of aircraft safety and organizational cost, respectively.

Introduction

Scheduled maintenance on an aircraft is typically associated with work cards, the purpose of which is to instruct vendors and their personnel on the specific areas requiring maintenance at a particular time. During the subsequent walk-around on the shop floor, the quality assurance representative checks the areas noted on these work cards, inspecting for maintenance errors and rating vendor performance based on the acceptance or the rejection of the work performed. If the quality assurance representative discovers a particular maintenance task mandated by the work card, which is not done to satisfaction, the vendor is instructed to correct and/or complete it before the aircraft leaves the facility. Each

---

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

Return to Table of Contents

time the task mandated by the work card is rejected by the quality assurance representative; an open-ended comment is documented.

In addition, the quality assurance representative may discover aspects of the aircraft requiring attention, which are not part of the scheduled surveillance. In this case, the maintenance issue is resolved during the surveillance event. If the finding is critical, during the next scheduled maintenance event, this aspect may be re-examined by both the vendor and the quality assurance representative. The typical surveillance response either accepts or rejects the quality of maintenance at the vendor location. The response to almost all rejects issued may also be associated with open-ended comments. These responses are vital sources of documented maintenance observations, carefully watched until the safety of the aircraft requires their resolution.

Given the importance of the open-ended comments made in the quality assurance process, it is critical to capture all open-ended response data in addition to the attribute data acquired from the forced responses. The quality assurance data to be analyzed by the web-based surveillance and auditing tool (WebSAT) are both qualitative and quantitative. WebSAT proposes to capture both types of information. The open-ended responses require interpretation to ensure their appropriate application in the maintenance/inspection process; therefore, this data must be associated with the appropriate performance measures of the maintenance process, reaching the personnel responsible for specific work functions. These performance measures indicate the impact of maintenance findings on the safety of an aircraft and the organization.

This data reduction is required at two levels. The first is the effectiveness of vendor performance in surveillance and technical audits, and the effectiveness of in-house processes measured by internal audits. This first level, referred to as Tier-1 in the proposed research, should indicate the impact of surveillance and audit findings on an aircraft. The second level, designated as Tier-2, involves the maintenance data captured by quality assurance representatives, auditors, and reflected in terms of cost implications at a managerial level.

The proposed research attempts to transform qualitative observations of vendor task performance into useful measures of aircraft safety and organizational cost. To accomplish such a transformation, it is important to assign an appropriate performance measure to qualitative observations.

This research involved quality assurance maintenance data generated during surveillance at the aircraft maintenance vendor locations of our industry partner, and the data generated during audits on internal processes and of processes at vendor location. The work functions associated with the quality assurance maintenance data are surveillance, technical audits, and internal audits. Surveillance is the day-to-day oversight and evaluation of the work contracted to an airframe substantial maintenance vendor to determine the level of compliance with our industry partner's Continuous Airworthiness Maintenance Program (CAMP) and General Maintenance Manual (GMM). The Quality Assurance Representative

(QAR), stationed at a vendor location, schedules surveillance of an incoming aircraft. The specific task to be performed on an aircraft is available on a work card. Depending on the quality of work performed by the vendor, a work card is either accepted or rejected by the QAR. Audits on the other hand involve the evaluation of processes. Internal audits involve evaluation of processes within the organization (our industry partner), and Technical audits involve evaluation of processes at the vendor location. An auditor has a checklist of either questions, which have a forced response of Yes or No. If the processes comply with standards, 'Yes' is associated with a checklist question, if not, 'No' is associated with the checklist question. Each time a QAR or an auditor come across discrepancies and 'Reject' the quality of work of a vendor, or associate 'No' with an audit checklist question, respectively, they may also document information associated with the discrepancy.

A strategy could be adopted where each documented open-ended response for a discrepancy is categorized into a performance measure by the QAR or an auditor. This is a straightforward way to transform qualitative observations into measures of aircraft safety and cost to the organization. This strategy has a drawback. The QARs and auditors each have their own interpretation of the documented discrepancies. If the transformation of the open-ended responses is done by, the QARs and the auditors there will be several similar performance measures within WebSAT. There is also a possibility that several of the measures will have similar meanings, creating redundancy. The managers warned us about this problem, and expressed the desire of having established measures for QARs and auditors to select. Hence, a better strategy needs to be adopted to transform qualitative data into meaningful performance measures.

Literature Review

Maintenance data could identify potential problem areas that may be indicators of aircraft safety. Maintenance data are both qualitative and quantitative. Quantitative models can predict system performance, whereas qualitative models may address important concepts that are not easy to quantify (Wang & Hwang, 2004).

Patankar and Taylor (2003) conducted a study utilizing the descriptive data from self-reported maintenance errors that are available through the Aviation Safety Reporting System (ASRS). They reported that the maintenance data provided a means to understand the relationship between causal factors that contribute to maintenance errors and the effect of these maintenance errors. Holmgren (2005) conducted a study to identify maintenance related losses and their causes. The study was conducted on the impact of maintenance related losses for Swedish railways, but Holmgren stated that such a strategy is of value to all concerned with transport and safety.

The Safety Performance Analysis System (SPAS) contains the Service Difficulty Reporting (SDR) system as a data source to identify abnormal and potentially unsafe conditions in an aircraft or aircraft components/equipment. A higher number of SDRs suggests a greater possibility of maintenance problems (Shyur, Luxhoj, & Williams, 1996). In their research on SPAS, Shyur et al. worked with performance indicators, the tracking of which helped identify unfavorable trends. Once performance indicators are identified, it is important to identify unfavorable trends from maintenance data.

Weckman, Shell, and Marvel (2001) developed a forecasting methodology based on actual field data for repairable systems in the aviation industry. Prediction models to determine aircraft maintenance errors can be based on multiple regression models. Multiple regression is a general statistical technique to analyze the relationship between a single dependent (predicted) variable and several independent (predictor) variables (Shyur, Luxhoj, & Williams, 1996). An effective maintenance classification and prediction method can help ensure a safe and operational aircraft. Appropriate data interpretation approaches can be utilized to find relationships between aircraft maintenance data and unsafe conditions of an aircraft (Brence & Brown, 2002). The discovery, collection, classification, and understanding of events for the purposes of developing specific error reduction strategies were recommended by Adams and Kirwan. (1997).

The computerized support system for analyzing human errors (COSFAH) is a unique approach to support the analyst during analysis. Yoon and Kim (1996) stated that the analysis reveals the cognitive causations among activities and system state, which in turn allow COSFAH to capture and provide rich insights into the causes of the errors.

Quantitative data are also a rich source of knowledge on aviation maintenance performance. It is critical to find hidden patterns in data sets to classify maintenance error. Thus, data mining is an aspect to be considered for a classification scheme. Data mining theories or techniques can be categorized into two categories. The first is the neural network and regression approach, which creates a model based on a training data set. The second category involves machine-learning algorithms generating a number of models in the form of decision rules. Data mining approaches have been successfully applied to industrial data analysis to derive useful and comprehensive knowledge (Han & Kamber, 2001).

Statistical language models have been found to be useful in encoding knowledge, making linguistic information available to data analysis systems. A review of the literature suggests that the most effective tool for this study should capture topic-related dependencies. It should assign an appropriate category to a certain type/kind of information, thereby translating existing linguistic knowledge into more useful responses through semantic information processing (Landauer, Foltz, & Laham, 1998). This type of approach appears to be critical to the current study involving the Web-based surveillance and auditing tool (WebSAT), as it allows the transformation of open-ended comments to meaningful maintenance performance measures. This is important because the open-ended responses of the quality assurance representatives and auditors track maintenance issues until they require action.

Multidimensional scaling (MDS) is a set of mathematical techniques enabling the hidden structure of databases to be uncovered (Kruskal & Wish, 1983). This class of techniques uses proximities among any kinds of objects as input, with proximity being defined as the number indicating how similar or how different two objects are, or are perceived to be. The resulting output is a spatial representation consisting of a geometric configuration of points that reflect the hidden structure of the data.

One of the attractive features of MDS is that it is not necessary to prejudge the structure of the databases (Rosenberg, Nelson, & Vivekananthan, 1968). The identification of the underlying patterns is the purpose of using a technique such as MDS.

In the current research to identify the categories for the cost to the organization, it seems important to incorporate the expertise of the managers. The auditors and quality assurance representatives need to be kept abreast of what the managers intend on achieving with the immense amount of maintenance data. A statistical technique such as MDS does not seem practical to identify performance measures of cost to the organization, since, in the case of our industry partner, there are only four managers, and the outcome of such a strategy would not be statistically significant. The Scandinavian school of thought of Participatory Design offers techniques in which users can directly influence the outcome of key design issues. Such a strategy could be useful to finalize the categories of cost to the organization.

The method of Participatory Design (PD) revolves around the concept that users and designers have equal significance in the development of a product. In conventional user-centered design practice, the users are often a part of the development process via user requirements gathering through interviews, questionnaires, observations, and evaluation sessions. The users typically no not participate directly in the decision making process.

In PD, the user(s) are involved with the decision making directly, as they are part of the design team. The team of users and designers is referred to as a multidisciplinary team. The study of PD has been an active research field for several decades, an indication that user involvement has an influence on the decision making process and that it generates insight and knowledge (Luck, 2003). Researchers have generally cited two reasons for adopting a PD approach: first being the fact that the study focuses on the verbal exchange of design ideas that is required during the early stages of the design process. The process is iterative and a better understanding emerges because of verbal exchange of ideas. The second reason hinges on the ideology of PD, which supports the participation of the users in the decision-making process (Luck).

The data collection process is primarily through semi-structured interviews to gather information from respondents. The rich information, which is produced, provides invaluable information for the design team (Luck, 2003). The key to this approach is that there is no right answer. Each response is supposed to be the user's personal response and experience. The analysis of the interview data deals with content analysis (Luck). This requires the researcher to be familiar with the data and to make objective inferences about characteristics within the text data. The textual data may have general user needs and may range to very specific user needs.

Strategies such as PICTIVE (Plastic Interface for Collaborative Technology Initiatives through Video Exploration) have been successfully implemented during a development process. This technique was introduced at Bellcore in 1990 within the framework of PD (Muller, 1992). This strategy has been successfully implemented to allow the direct participation of users who may not be computer literate in the software design process. The basic idea was to capture ideas expressed during the PICTIVE sessions on video, for later interpretation and implementation in the design process. The expertise of the researcher, the commitment of the users and stakeholders to a successful product, and workplace democracy would be key requirements for the success of PD in the workplace (Muller).

PD has questioned traditional work practices, especially in an era when businesses are abandoning their traditional practices to embrace new ideas (Muller, Wildman, & White, 1993). PD researchers have been exploring the appropriate conditions for user participation in the design and introduction of computer-based systems at work (Kensing & Blomberg, 1998). According to Kensing and Blomberg, PD projects have varied with respect to how and why workers have participated. In some cases, worker participation is limited to providing designers with access to workers' skills and experiences, while in others, workers participate because their interests in the design outcome are critical. Cooperative Experimental Systems Development (CESD) is characterized by user involvement throughout the development process (Gronbaek, Kyng, & Mogensen, 1997), while Contextual Design focuses on early design activities (Beyer & Holtzblatt, 1997).

Blomberg and Henderson (1990) have defined three tenets of the PD approach that should influence the interaction between developers and users of computer-based systems: 1) the goal should be to improve the work life of the users; 2) the orientation should be toward collaborative development; and 3) the process should be iterative.

This research, then, proposes to apply multidimensional scaling (MDS) to the open-ended responses of quality assurance representatives and auditors, to transform qualitative maintenance data to established performance measures implying the impact of maintenance findings on the safety of the aircraft. Both Taguchi philosophy and robust design methodology support the importance of this data reduction, emphasizing that effective data analysis depends directly on the appropriateness with which a system, in this case WebSAT, reduces all the associated work function information (Cho, Kim, Kimbler, & Phillip, 2000; Kim & Cho, 2000a; Kim & Cho, 2000b; Taguchi, 1986). It is especially critical that textual data associated with open-ended responses be analyzed accurately, as these responses indicate aspects of an aircraft requiring future maintenance attention. As such, it is important to identify which attribute of a work function is affected.

In the third phase of the research, PD methodologies will be applied to enable the managers of the quality assurance department at our industry partner, to form appropriate performance measures against which maintenance data may be categorized in terms of cost to the organization.

Problem Statement 1

Each time a QAR or an auditor observes a reject or a “No,” respectively, an open-ended response is documented. The impact on the safety of the aircraft is not indicated by the open-ended comment. Each time a reject or a No is documented, an associated aircraft level impact (ALI) needs to be associated with the finding. The first phase of the research, then, proposes to apply MDS to categorize the open-ended maintenance data into useful performance measures. Each open-ended response indicates a possible risk to the aircraft and will be associated with an ALI.

Problem Statement 2

The users of WebSAT at the managerial level are concerned with the implications of discrepancies observed by quality assurance representatives and auditors in terms of a potentially different set of managerial-level categories (Organizational Categories). There are not enough managers to allow a statistical technique such as MDS to be appropriately used to establish this set of categories. This research then proposes to use Participatory Design to finalize the set of Organizational Categories.

Multidimensional Scaling and Card Sorting

Research requirements

Each time a reject is observed by a quality assurance representative, an open-ended comment is documented. A categorization of this response is then made in terms of its effect on aircraft safety. Such a categorization is defined by the surveillance personnel as the effect of the rejection on the safety of an aircraft. Currently, only the quality assurance representatives categorize a reject into an ‘effect’ category. The categories for ‘effect’ are general workmanship, hangar safety, safety immediate, lubrication, regulatory compliance, safety personnel, and operational. Although the auditors do not currently categorize audit findings, the open-ended responses, which they document, could also be utilized to identify the effect of the finding on the safety of an aircraft. The open-ended responses from previous surveillance findings and audits will be presented to quality assurance representatives and auditors for this research. They will be asked to categorize the findings into performance metrics indicating the effect of a finding on aircraft safety. This research then proposes to use MDS to utilize these open-ended responses to establish categories for the safety of an aircraft.

Multidimensional Scaling

Multidimensional Scaling (MDS) refers to a class of techniques. MDS uses proximities among any kinds of objects as input. Proximity is a number that indicates how similar or dissimilar two objects are. The output is a spatial representation of a geometric configuration of points (Kruskal & Wish, 1983). Each point in the configuration corresponds to one object. The configuration represents the structure in the data. This makes the data easier to understand. The larger the dissimilarity between two objects, the further apart they will be in a spatial configuration.

A common procedure to obtain proximity data is to ask people to directly judge the closeness of the stimulus objects (Kruskal & Wish, 1983). In order to discover dimensions and not impose them, the attributes on which the stimuli are to be judged are not specified. A simple method for large stimulus sets (approximately 50 to 100 objects) is to have subjects sort or cluster the stimuli according to perceived similarity (Rosenberg et al., 1968).

The subjects in this study were asked to place the stimuli into exclusive and exhaustive categories. Thus, stimuli in the same category are more similar to each other than to those in other categories. In MDS analysis, greater distance between objects reflects less interaction between the associated entities (Jones & Young, 1972)

The data for analysis could pertain to some collection of objects (Kruskal & Wish, 1983). The stimuli could be either real or conceptual. The objects can be primarily indexed by the letter i and secondarily by j, and i and j run from one to i if there are i objects. The proximity-associating object i and j can be represented by δij. There is no effective difference between δij and δji. There is no effective value associated with δii. The distance between points plays an important role in MDS (Kruskal & Wish). The distance between two points x_i and x_j can be represented by d (x_i , x_j), and this is usually simplified further to d_{ij} . The distance is always Euclidean distance, unless stated otherwise.

For any set of data (Kruskal & Wish, 1983), the objective function yields a single number, which shows how well or how poorly the data fit the configuration. Thus, f (δij) = d_{ij} , where f is some specified type of function. The discrepancy between f (δij) and d_{ij} is then f (δij) - d_{ij} .

Kruskal and Wish (1983) stated that:
If f (δij) = d_{ij} , then we have exact equality.

The “f-stress” is a measure of the configuration (Kruskal & Wish, 1983). The larger the f-stress, the worse the configuration is. The value for f-stress is ≥ 0, since we account for squared distances. If we have exact equality, then f-stress = 0.

A rough rule of thumb is that there should be at least twice as many stimulus pairs as parameters to be estimated. This assures an adequate degree of statistical stability (Kruskal & Wish, 1983).

The MDS technique appears to be suited to the current problem of understanding and describing the multidimensional structure of aircraft maintenance data. The application of MDS for the study will start with the selection of a set of objects, a large and diverse set of open-ended maintenance data documented by quality assurance representatives and auditors. The basic datum required for MDS (Rosenberg et al., 1968) is a number for each pair of objects in the selected set of objects reflecting how closely the two objects are related to each other. What MDS then does is provide a geometric representation of the set of these objects (or responses in the case of this research) so that the inter item distance in space corresponds to the empirical measure of psychological relatedness (Rosenberg et al. 1968).

Research phases

There are three phases to identify and validate the categories of impact to aircraft safety. These are as follows: 1) Card-sorting technique to understand proximity distances between audit and surveillance findings; 2) Validate dimensions or categories identified in the first phase by rating the relationship between the findings and the dimensions (or categories) identified; and 3) Validate the utility of the dimensions (or categories).

Card Sorting Technique

A card sorting technique was used to generate a set of categories of the documented audit and surveillance findings. The results from this study were used to establish an initial set of categories that indicate the effect of maintenance error on the safety of an aircraft. The audit and surveillance findings, which were used for this study, were documented findings by technical auditors, internal auditors, and quality assurance representatives from the surveillance department for the fiscal years 2003 to 2005, 2002 to 2006, and 2003 to 2005, respectively. The researcher selected unique findings from the entire data set for the work functions of technical audits, internal audits, and surveillance. Three-hundred and four responses were selected for this study. The researcher and three other graduate students from the Department of Industrial Engineering, Clemson University, who were aware of the card sorting technique, did a pilot study involving the sorting task. This was done to determine an approximate time range, which would be required, to do the card-sorting task for such a large and diverse set of findings. The average time was approximately 90 minutes. The time for the participants for the actual study was set at 120 minutes. Since the task was designed to be a self-paced task, 120 minutes was only a recommended time limit. The participants could take more or less time for the task, than the recommended time limit.

Participants

Fifteen participants from our industry partner for this research were asked to participate in the card sorting study. Seven quality assurance representatives from the surveillance department, five internal auditors, two technical auditors, and a manager from the technical audit department were involved in this study.

Task

The participants were given the task of categorizing the responses into clusters of similar responses. There was no restriction on the number of categories the participants could generate. The subjects were asked to keep the responses on the cards face up and to place a slip of paper next to each cluster of responses indicating an apt name for the cluster of similar responses. They were allowed to change their assignments of responses to a new category at any time before they completed the task. Even though the informational letter provided to the subjects indicated that the task would take approximately 2 hours, the researcher stressed the fact that the task was self-paced, and the participants could take more time. This never happened during the study, since the longest time taken by any participant was approximately 115 minutes. The range of times for this task was between 90 and 115 minutes, while the mean time for the task was 107.33 minutes.

Apparatus and Settings

The card sorting technique was conducted at the industry partner headquarters in Memphis, Tennessee, and at aircraft vendor locations at Greensboro, North

Carolina, and Mobile, Alabama. Three hundred and four cards representing audit and surveillance findings were presented to the participants.

Procedure

Participants were invited to participate in the card sorting study through phone calls. An Informational Letter was presented to the participants. The instructions for the study were documented in this letter. The instructions were also read aloud to the participants. The participants were allowed to ask questions regarding the study, and only when each participant knew exactly what needed to be done, did they proceed with the study.

Results

The number of categories used by the fifteen participants ranged from three to fifteen with a mode of five categories. Kruskal and Wish (1983) suggested that a stress of 5% is 'good', while 10% is 'fair'. The plot in Figure 1 suggests that the stress value improves from two dimensions onwards. For four dimensions, the stress value is 0.08181 and for five dimensions, the stress value is 0.07019, or approximately 7%. Even though the stress values for six and seven dimensions are lower than that for five dimensions, the loadings or association of responses on the sixth and seventh dimension are largely insignificant..

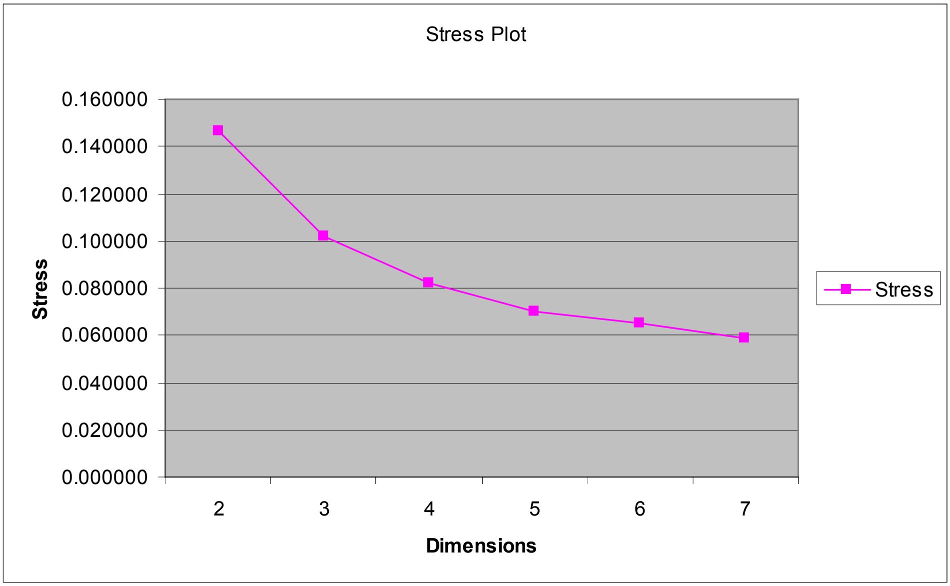


Figure 1. Stress plot for identifying the number of dimensions

According to Kruskal and Wish (1983), the stress value should stabilize and the cut off point where the value stabilizes will be the number of dimensions that account for the most items. The dimensions or categories at this stage of the research were finalized at five.

Identified Dimensions

The ten responses with the highest loadings on the five dimensions were studied carefully. The selection criteria for the responses are explained in the *Linear Regression* subsection. Each of the responses was then associated with the names of clusters which the participants provided during the study. The five dimensions are as follows: Safety, Regulatory Compliance, Procedure / Paperwork Inadequacy, Operational, and Housekeeping and Storage.

Typical findings for safety would be: 1) Pulled circuit breakers without "Lock Out Tags" in cockpit of N673FE, and 2) Found fire extinguisher #EE 4328 past due inspection on lift truck # 10165.

Findings for regulatory compliance would be as follows: 1) Several Aircraft Maintenance Technician (AMT) training records contained multiple and/or outdated copies of Measures of Performance (MOPs) and several AMT records had incomplete MOPs past the 60 day time limit for completion, and 2) Internal audits are not being conducted in accordance with the frequency outlined in the Base Maintenance Desktop Procedure Manual (DTPM).

Findings for procedure and paperwork inadequacy would include findings such as 1) Investigation revealed that stamp control as per the Base Maintenance DTPM was not being followed consistently, and 2) Found engine received and released for service with incorrect or missing component airworthiness approval documentation.

Typical operations findings include responses such as: 1) Several ball-mats were found without serviceability documentation in two separate locations, and 2) Found numerous main A/C batteries on pallets in the warehouse area without paperwork attached to identify them.

Housekeeping and storage includes findings such as 1) Scrap trailers full and outside area overflowing, and 2) Found box with lose hardware at nose of aircraft.

Rating Responses on Identified Dimensions

This phase of the research was important to establish a set of stable dimensions with empirical data. At the end of the *Identified Dimensions* subsection, five dimensions were uncovered in the configuration space. The loading values for five dimensions indicate that the loading values for the fifth dimension are comparatively low. This phase will then generate empirical data to establish a final set of dimensions.

Linear regression to interpret multidimensional solutions

The directions in the MDS configuration have interesting interpretations. The positions in the configuration may be associated with characteristics of the items that were scaled (Kruskal & Wish, 1983). One of the key expectations of MDS is to understand which characteristics are more important than the others are. Linear regression is the most common way to identify these characteristics. Kruskal and Wish explained the utility of linear regression. Assume there is some variable associated with items that has a systematic relationship to position in the configuration. One way to establish this relationship is to perform linear regression using this variable as the dependent variable and the coordinates of the configuration as the independent variables. Generally participants in a study are asked to rate the association of a response with the identified dimensions. The mean value of the ratings forms the dependent variable. The ordinary multiple R provides a quantitative estimate of the degree to which an item actually corresponds to a dimension in the space configuration (Rosenberg et al., 1968).

Linear regression

The mean rating of each response on each dimension is the dependent variable. There are five dependent variables: the mean ratings on each of the five identified dimensions. The coordinates of the configuration (or loadings) for the five dimensions are the independent variables. Five regression studies were conducted, one for each of the dimensions.

Regression analysis is a statistical technique that allows the assessment of the relationship between one dependent variable (DV) and several independent variables (IV). The regression equation takes the following form:

$Y' = A + B_1X_1 + B_2X_2 + ... + B_kX_k$, where Y' is the predicted value on the dependent variable, A is the Y intercept (the value of Y when all the X values are zero), the X (i) s represent the independent variables (of which there are k in the above case), and the B (i) s are the coefficients assigned to each of the independent variables during regression. The goal of regression analysis is to obtain a set of B values, known as regression coefficients, for the IVs that bring the Y values predicted from the equation close to the Y values obtained by measurement (Tabachnick & Fidell, 2001).

In this stage of the study, the responses used for clustering were used to obtain regression coefficients for the association between the independent and dependent variables. Three hundred and four maintenance findings were used in the initial study to categorize these findings. The participants for the earlier study and for the regression study are auditors and quality assurance representatives. It was difficult to secure the participation of these stakeholders for extended periods. Thus, a reduced number of findings were used for this next study. Green (1991) provides a discussion regarding the sample size requirement. The simplest rule of thumb is that the sample size, $N \geq 50 + 8m$, where m is the number of IVs, to test the multiple correlation and that $N \geq 104 + m$ to test individual predictors. Green (1991) offers a more complex rule of thumb which accounts for the effect size: N ≥ (8/f-sq.) + (m-1), where f-sq. = 0.01, 0.15, and 0.35 for small, medium, and large effects, respectively.

For five dimensions, and a medium effect size, the sample size should be approximately fifty-seven. It was then decided to select an equal and appropriate number of responses from each of the identified dimensions. It was decided that the highest loading values on the negative and positive axis would be accounted for. The minimum acceptable loading value for a response was set at 0.7. This created problems since for the third, fourth, and fifth dimensions, the loading values were much lower than 0.7. The acceptable threshold was then lowered to 0.5. The five highest loading values for the fifth dimension range between -0.521 and -0.318 on the negative axis and between 0.453 and 0.29 on the positive axis. Beyond this, the loading values have a dramatic drop much below the acceptable loading values. It was then decided to consider five responses each on the negative and positive axis for each dimension.

By considering ten responses for each dimension, the sample size for this study is fifty. This is close to the fifty-seven which Green (1991) proposed. There are three responses that have similar loading values on two dimensions, hence there were forty-seven responses used in the study.

Rating of Responses

Each subject in this phase was given a list of forty-seven responses. The subjects were asked to rate these responses on a scale of one to seven. Five such scales were used to identify the association of responses with the identified dimensions. The responses used in this stage were surveillance and auditing findings. The five highest loading values on both the positive and negative scale for each dimension were included in the response set. The higher the negative or positive value of a loading for a response on a dimension, the more strongly this dimension is associated with the response.

Participants

All participants were personnel from the quality assurance department of our industry partner. There were fifteen participants in this study. There were five technical auditors, five internal auditors, and five quality assurance representatives from the surveillance department.

Task

Each participant was given a set of 47 responses on a word document. They were then asked to indicate on a scale of one to seven, the association of these responses with each of the five dimensions identified. There were five rating tasks, one for each identified dimension.

Apparatus and Settings

The rating study was conducted at the industry partner headquarters in Memphis, Tennessee, and aircraft vendor locations at Greensboro, North Carolina, and Mobile, Alabama. A word document was provided to each participant for each of the five rating schemes.

Procedure

Participants were invited to participate in the rating of responses study through phone calls. An Informational Letter was presented to the participants. The instructions for the study are documented in this letter. The instructions were read aloud to the participants. The participants were allowed to ask questions regarding the study, and only when each participant knew exactly what needed to be done, did they proceed with the study.

Results for regression analysis

The results were interpreted in two stages. In the first stage, the data were analyzed for the study as it happened, i.e., for the common pool of participants from the quality assurance department. In the second stage, the dimensions were verified for individual work function participants. This was done to make concrete sense of identified dimensions.

There were three statistical estimates used to interpret the results. These were the p-value (p), regression coefficient (R-sq.), and the un-standardized coefficients (B). The un-standardized coefficients of the regression analysis are the most important estimate indicating the similarity or dissimilarity of identified dimensions.

In the initial stages, qualitative aircraft maintenance data were used to identify performance measures (dimensions for MDS) indicating the safety of an aircraft. There were five dimensions that were identified. These dimensions were interpreted as safety, regulatory compliance, procedures and paperwork inadequacy, operations, and housekeeping and storage.

In the next stage, linear regression was used to verify the association between the identified dimensions from MDS and how the quality assurance personnel interpret these as performance measures to quantify qualitative aircraft maintenance data. The loading values of the dimensions from MDS are the independent variables for this study, while the rating value obtained for each open-ended response is the dependent variable.

Safety, regulatory compliance, procedures and paperwork inadequacy, operations, and housekeeping and storage are referred to as dimension 1 through dimension 5, respectively.

Regression analysis for quality assurance personnel

For safety (rating associated with dimension 1), the p-value is 0.012. This is significant. The R-sq. is 0.29. The B value for procedures and paperwork inadequacy (loading on dimension 3 from MDS) is -2.79. What this signifies is that dimension 1 and dimension 3 are opposite to each other.

At this point, it is critical to explain the significance of the statement that dimension 1 and dimension 3 are opposite to each other. In the context stated above, the participants felt that certain responses can be explained more in the context of ‘safety’ (dimension 1), and have less to do with ‘procedures and paperwork inadequacy’ (dimension 3). A B value of negative also indicates the fact that dimension 1 and dimension 3 are perceived to be different.

This can further be interpreted as follows: as the loading on dimension 3 (procedures and paperwork inadequacy) is increased by one unit, the inferred rating for dimension 1 (safety, which is the predicted variable or the dependent variable) goes down by -2.79 units.

The p-value for regulatory compliance (rating associated with dimension 2) is insignificant at 0.156. The R-sq. is 0.172. None of the B values are interpreted to anything of significance.

For procedures and paperwork inadequacy (rating associated with dimension 3), the p-value is 0.002, which is very significant. The R-sq. is 0.364. The B value for dimension 3 is 2.072, which interprets to the fact that dimension 3 was found to be significant and meaningful to the quality assurance personnel. For each unit increase in the loading on dimension 3, the inferred increase in the rating goes up by 2.072 units for procedures and paperwork inadequacy.

The p-value for operations (rating associated with dimension 4) is 0.607, which is not significant. The R-sq. is 0.081. None of the B values are significant. The B value for dimension 5 is -1.004, indicating that the participants did interpret dimension 4 and dimension 5 as opposite measures of quantitative maintenance data. This can be interpreted further: for each unit increase of loading on dimension 5 (housekeeping and storage), the inferred rating goes down by -1.004 units on dimension 4 (operations).

The p-value for housekeeping and storage (rating associated with dimension 5) is 0.23, which is not significant. The R-sq. is 0.15. The B value for operations (dimension 4) is -2.396, indicating that the participants found dimension 4 and dimension 5 to be opposite to each other. Further, for each unit increase in the loading values for dimension 4, the inferred rating on housekeeping and storage (dimension 5) goes down by -2.396 units.

Regression analysis for Internal audits personnel

The p-value for safety (rating associated with dimension 1) is 0.009, which is significant. The R-sq. is 0.302. The B value for dimension 3 is -2.929, indicating that the internal audit personnel interpret dimensions 1 and 3 to be opposite to each other. For each unit increase in the loading for dimension 3 (procedures and paperwork inadequacy), the inferred rating value for safety (dimension 1) goes down by 2.929 units. The B value for dimension 3 has increased compared to the observed B value for dimension 3 when the participant data were pooled together, indicating that internal audit personnel interpret dimension 1 and 3 to be opposite to each other more significantly than what was indicated in the pooled data set.

The p-value for regulatory compliance (rating associated with dimension 2) is insignificant with a value of 0.262. The R-sq. is 0.142. None of the B values are of any significance.

The p-value for procedures and paperwork inadequacy (rating associated with dimension 3) is significant at zero (0 < 0.05, where alpha is 0.05). The R-sq. is high at 0.53. The B value for dimension 3 is very high at 3.762, indicating that the internal audit personnel found dimension 3 indicative of procedures and paperwork inadequacy. For each unit increase in the loadings for dimension 3 (procedures and paperwork inadequacy), the inferred increase in the rating for procedures and paperwork inadequacy (dimension 3) is 3.762 units. The B value for dimension 3 increases compared to the observed B value for dimension 3 for the analysis of pooled data for quality assurance personnel.

The p-value for operations (rating associated with dimension 4) is insignificant at 0.142. The R-sq. is 0.177. None of the B values indicate anything significant.

The p-value for housekeeping and storage (rating associated with dimension 5) is insignificant at 0.211. The R-sq. is 0.155. The B value for dimension 4 is -3.108, indicating that internal audits personnel observed dimension 4 and dimension 5 to be significantly different. Further, for each unit increase in the loading for dimension 4, the inferred decrease in the rating value is 3.108 units. The B value for dimension 4 indicates that internal audits personnel find dimension 4 and 5 to be different with a higher degree of certainty than the quality assurance personnel do.

Regression analysis for Surveillance personnel

The p-value for safety (rating associated with dimension 1) is significant at 0.04. The R-sq. is 0.24. The B value for dimension 3 is -2.816, which is marginally higher than that observed for the data associated with the quality assurance personnel, but lower than what was observed for the internal audits personnel. The B value for dimension 3 indicates that the surveillance personnel observe dimension 1 and 3 to be opposite indicators of qualitative aircraft maintenance data. Further, for each unit increase in the loading value for dimension 3 (procedures and paperwork inadequacy), the decrease in the inferred rating for safety (dimension 1) is by 2.816 units.

The p-value for regulatory compliance (rating associated with dimension 2) is insignificant at 0.163. The R-sq. is 0.169. None of the B values indicate anything significant.

The p-value for procedures and paperwork inadequacy (rating associated with dimension 3) is significant at 0.047. The R-sq. is 0.232. The B value for dimension 3 is 1.444, which is less significant than that for the internal audits personnel. For a unit increase in the loading value for dimension 3 (procedures and paperwork inadequacy), the inferred increase in the rating for procedures and paperwork inadequacy is by 1.444 units.

The p-value for operations (rating associated with dimension 4) is insignificant at 0.329. The R-sq. is 0.127. The B value for dimension 4 (operations) is 1.383, indicating that surveillance personnel observed dimension 4 to be a true indicator of operations as a performance measure of qualitative aircraft maintenance data. For each unit increase in the loading value for dimension 4 (operations), the inferred increase in the rating for dimension 4 (operations) is 1.383 units.

The p-value for housekeeping and storage (rating associated with dimension 5) is significant at 0.04. The R-sq. is 0.24. The B value for dimension 4 is -1.030, indicating that the surveillance personnel observed dimensions 4 and 5 to be opposite to each other. For each unit increase in the loadings for dimension 4 (operations), the inferred decrease in the rating for housekeeping and storage (dimension 5) is by 1.03 units.

Regression analysis for Technical audits personnel

The p-value for safety (rating associated with dimension 1) is significant at 0.027. The R-sq. is 0.258. The B value for dimension 3 (procedures and paperwork inadequacy) is -2.573, indicating that the technical audits personnel observed dimension 1 and 3 to be opposite to each other. For each unit increase in the loading for dimension 3 (procedures and paperwork inadequacy), the inferred decrease in the rating for safety (dimension 1) is by -2.573 units. The B value for dimension 4 is 2.385, indicating that the technical audits personnel observed dimensions 3 and 4 to be significantly opposite to each other. For each unit increase in the loading value for dimension 4 (operations), the inferred increase in the rating for safety (dimension 1) is by 2.385 units.

The p-value for regulatory compliance (rating associated with dimension 2) is insignificant at 0.267. The R-sq. is 0.14. None of the B values indicate anything of significance.

The p-value for procedures and paperwork inadequacy (rating associated with diension 3) is insignificant at 0.587. The R-sq. is 0.084. The B values for dimensions 4 and 5 are -0.735 and 1.697, indicating that these two dimensions are observed to be opposite to each other. For each unit increase in loading for dimension 4 (operations) the inferred decrease in rating for procedures and paperwork, inadequacy (dimension 3) is by 0.735 units. For each unit increase in the loading for dimension 5 (housekeeping and storage), the inferred increase in the rating for procedures and paperwork inadequacy (dimension 3) is by 1.697 units.

The p-value for operations (rating associated with dimension 4) is insignificant at 0.307. The R-sq. is 0.132. The B value for dimension 4 (operations) is 1.007, indicating that this dimension is an indicator of the performance measure, operations, which is the identified dimension 4 in this case. Each unit increase in the loading values for operations (dimension 4), has an inferred increase on the rating for operations (dimension 4) by 1.007 units.

The p-value for housekeeping and storage (rating associated with dimension 5) is insignificant at 0.069. The R-sq. is 0.214. The B value for dimension 4 (operations) is -4.175, indicating that dimension 4 and 5 are observed to be opposite to each other by the technical audits personnel. For each unit increase in the loading for dimension 4 (operations), the inferred decrease in the rating for housekeeping and storage (dimension 5) is by 4.175 units.

Discussion

Results confirm that dimensions 1 (safety), 3 (procedures / paperwork inadequacy), 4 (operations), and 5 (housekeeping and storage) are inferred by various work function groups to be valid. Dimension 2 (regulatory compliance) surprisingly did not have significant p-values for any group.

Based on the loading values for five dimensions it is concluded that Regulatory Compliance (dimension 2) was better discriminated between items (open-ended qualitative maintenance data). This is explained further. Dimension 2 was more conclusive a dimension to explain the impact on aircraft safety, than were dimensions 3, 4, and 5. Within the inductive approach of MDS, dimension 2 meant something to the quality assurance representatives and auditors. They were able to discriminate between items on dimension 2.

The R-sq. describes the sample size effect and not the population effect. The effect sizes (R-sq. values) for this study were low. The B-values for this study were promising.

Thus, taking the loading values from MDS and linear regression analysis results into account, it is concluded that the five identified dimensions are interpreted to be useful and meaningful by the quality assurance personnel.

In the next stage of the research, Organizational Categories were identified to associate a documented surveillance or audit finding with performance measures of organizational cost. An initial unsuccessful card sorting study with quality assurance representatives and auditors, allowed the researcher to explore other options. The strategy was flawed because for performance measures at an organizational level, the quality assurance representatives and auditors were not a representative group. The researcher then adopted Participatory Design to identify the required performance measures.

Participatory Design for Organizational Categories

A requirement of WebSAT is that it transforms the open-ended responses associated with reject and ‘No’ into categories of system cost at the managerial level. No such categories currently exist for the quality assurance department. However, the managers of the quality assurance department have suggested that categories such as Human Factors, Economics, and Quality might be appropriate to describe cost implications at the managerial level. The research, then, proposes to present open-ended responses to managers in the Quality Assurance department and apply the strategy of Participatory Design to conduct a study with the managers to finalize a new set of managerial-level categories (Organizational Categories).

Participatory Design

Participatory design is a design strategy to include designers and users in a multidisciplinary team to allow a verbal exchange of ideas through semi-structured interviews to gather information from respondents. It is critical for the researcher to derive objective inferences from these sessions. The experiences of the users and their interests in the design outcome improve the work life of the users through collaborative development.

In this phase of the research, participatory design methodology will be utilized to establish Organizational Categories to categorize audit and surveillance findings.

Participants

Three participants from the quality assurance department were asked to participate in this study. One manager from the surveillance department and one manager each from the technical and internal audits department were a part of this study.

Setting

The participatory design study was conducted at the industry partner headquarters in Memphis, Tennessee. The researcher had a list of questions that were the basis for a semi-structured interview.

Procedure

The participants were asked for their consent to participate in the study prior to the start of the study. The managers were presented with two alternative categorization schemes, which the research team had developed through interactions, interviews, and studies conducted with the quality assurance personnel. The first set of categories includes performance metrics such as Human Factors, Quality, and Economic, which the managers had mentioned. The second set of categories was developed based on a card sorting study with auditors and quality assurance representatives. The results of this study revealed only one relevant dimension, at three levels. The auditors and quality assurance representatives understood the relevance of organizational categories, but indicated that they think of organizational categories with respect to cost only. Hence, the only dimension, which was uncovered, was cost, at the three levels of low, medium, and high.

A series of questions were asked which allowed the end users of the product to be a part of the final decision making process.

Problem statement description

The stakeholders of the project required the qualitative maintenance data to be interpreted to indicate the implications of aircraft maintenance findings on the organization. WebSAT will collect and analyze data associated with aircraft maintenance. The departments associated with aircraft maintenance are surveillance, technical audits, and internal audits. None of these departments have any kind of performance measures to indicate the implications of aircraft maintenance findings at an organizational level. The purpose of this research is to establish performance measures associated with the impact of aircraft maintenance findings on the organization.

Data collection process

The process of Participatory Design to identify the performance measures depends on interview data gathered from managers associated with the departments of surveillance, technical audits, and internal audits. These data illustrate the understanding and interpretations of managers, who directly influence the organizational level decision making process associated with aircraft maintenance findings.

The interviews were conducted with the managers of the quality assurance department to understand the expectations of the managers regarding performance measures at an organizational level. The number of managers interviewed was small, just three. One manager from each department took part in the study.

The interview was documented by the interviewer so that it could be reviewed at a later stage -- before the performance, measures for organizational categories were finalized. The documented notes from the interviews reflect an iterative decision-making process, where the managers helped the interviewer reach a final decision. The advantage of this study was the interaction with managers who were the primary stakeholders of this aspect of the project. A previous study with the auditors and quality assurance representatives to finalize performance measures for organizational categories did not work out. The study involved quality assurance representatives and auditors. The participants were given the task of categorizing maintenance findings into clusters of similar responses. Each cluster was associated with a category implying the impact of maintenance findings on the organization. The failure is attributed to the fact that the auditors and the quality assurance representatives were a non-representative group to uncover meaningful organizational categories.

Semi-structured interviews

A questions checklist was used to help the interviewer conduct the semi-structured interview. The checklist questions were developed based on experience in user-centered design and understanding of the problem at hand to finalize performance measures for organizational categories. The series of questions were aspects that needed to be discussed with the managers to understand their interpretation of the performance measures. Each question prompted comments from the managers, which furthered the discussion. The rich information this produced was valuable for the study and the researcher. It helped reach a final decision on organizational categories that would be used to indicate the implications of surveillance and audit findings at an organizational level.

The interviews were conducted using the same questions checklist for each manager. This allowed consistency across the managers interviewed but did not impose a structure on the responses of managers. The responses were spontaneous and unbounded. This approach generated rich and valuable information about the personal perceptions of the managers. The advantage of such an approach was that the managers were interviewed separately and hence they had the advantage to put across their ideas individually, without the influence of the other participants.

Each response was an idea that helped reach a final decision on an attribute of the product.

Interview data

Each interview varied in length and took 1-1.5 hours to complete. The interviews were scripted into a text document. The researcher made several inferences based on the data, which were generated. The researcher read the data several times to become familiar with the interview data. This helped the researcher identify the significant factors in determining the performance measures.

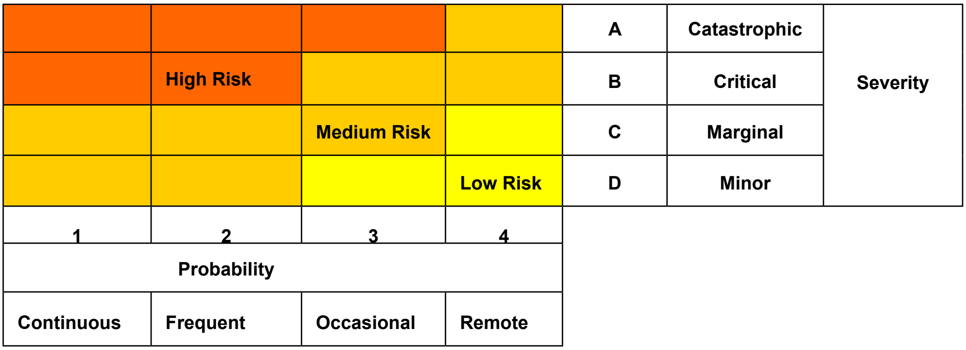
Interview with the Internal audits manager

The manager of the internal audits department was reminded of categories such as Human Factors, Economic, and Quality, which were recommended to the product development team by the quality assurance managers as appropriate categories to indicate the impact of surveillance and audit findings at an organizational level. The manager responded by saying that these categories were considered a good starting point to finalize worthwhile and meaningful organizational categories. These categories were recommended to the product development team, of which the researcher is a member, in the initial stages of the project. The manager felt that, now, these categories appeared vague to him. He mentioned that during the initial stages of the project these categories might have been mentioned to motivate the quality assurance department and the product development team to research the aspect of organizational impact associated with surveillance and audit findings. The manager felt, now, the mentioned categories did not make much sense to him and he offered his reasons.

According to the manager Human Factors, Economic, and Quality are very broad categories. The manager expressed his uncertainty about having any of these categories as established categories for organizational categories. According to the manager there are two major concerns regarding the mentioned categories. Firstly, each category appears to require sub-categories to make some concrete sense out of their implications. Secondly, the manager mentioned that categories such as Economic and Quality work two ways; they could either imply a root cause category indicating what caused an audit finding, or they could be used to imply the consequence of the audit finding at an organizational level. He said that this would create more uncertainty in the minds of the auditors while they were documenting audit findings and categorizing an appropriate organizational category associated with the audit finding.

The manager expressed additional concerns. He said that Human Factors is a gigantic category that will create more problems in the mind of the auditor, than help resolve anything.

The organizational category was referred to by the manager as consequences after a condition is found in an audit. The manager informed the interviewer about the risk management matrix used by the internal audit management group (Figure 2).



				A	Catastrophic	Severity
	High Risk			B	Critical	
		Medium Risk		C	Marginal	
			Low Risk	D	Minor	
1	2	3	4			
Probability						
Continuous	Frequent	Occasional	Remote			

Figure 2. Risk Management Matrix used by the Internal Audits department

The Risk Management Matrix (RMM) consists of four severity categories and four likelihood categories or probabilities. Each audit finding is associated with a risk level depending on the severity of the finding, and the likelihood associated with the occurrence of the finding. The management currently uses this matrix to indicate a ‘risk’ associated with an audit finding. The combination of a severity and the probability of occurrence help the management identify the risk level of an audit finding. Further, each severity is associated with three factors, which help define the extent of the problem. The three factors are performance, people, and machine.

The ‘performance’ associated with each severity defines the typical situation which may arise due to an audit finding (Figure 2). For example, a ‘minor’ severity occurs when the audit finding has a minimal system consequence.

The RMM also has a factor termed ‘people’, which is associated with the severity (Figure 2). This helps the auditor or the manager identify the risk to people when the audit finding occurs. For example, if personnel required first aid due to an injury, the severity associated would be ‘marginal’.

The third factor is ‘machine’. This indicates the dollar value assigned to the loss sustained by the organization due to an audit finding (Figure 2). A ‘critical’ severity is associated with an audit finding that accounted for losses worth more than $100,000.

RMM has definitions for the probability in terms of ‘occurrence’ and ‘rate’ (Figure 2).

The manager said that the internal auditors have been using the RMM for a while, and the strategy is working for their department to identify the ‘risk’ associated with audit findings at an organizational level.

The manager informed the interviewer that an internal auditor should indicate the risk associated with audit findings. The manager should be able to edit the auditors’ classification of risk associated with audit findings.

The interviewer informed the manager about a research study conducted with quality assurance representatives and auditors to identify suitable performance measures for organizational categories. The study indicated that the auditors and quality assurance representatives categorized surveillance and audit findings in terms of high cost, medium cost, low cost, and no impact to indicate implications of audit and surveillance findings at an organizational level.

The manager was not impressed with the categories and informed the interviewer that he would much rather have the auditors use the RMM.

Interview with the Technical audits manager

The interviewer informed the manager about two sets of categories which were uncovered by the WebSAT research team. The first set of categories consist of factors such as Human Factors, Quality, and Economic, which were recommended to the product development team by the quality assurance managers during the initial stages of the project.

The manager was informed that the second set of categories consists of cost at three levels of low, medium, and high. These categories were uncovered during an unsuccessful attempt to research valid performance measures to indicate the organizational impact of surveillance and audit findings. The study was conducted with quality assurance representatives from the surveillance department and auditors from the departments of internal and technical audits. The manager was informed that some auditors and quality assurance representatives categorized some findings as having no impact to the organization.

The manager was not confident about either set of categories and offered his explanation.

The manager indicated that factors such as Human Factors, Quality, and Economic would indicate the ‘cause’ of a finding. Human factors are a cause and include a variety of factors such as time, errors, equipment failure, avoidance of work, and lack of knowledge or understanding of a process. Inadequate procedures would also be a cause category and would involve findings caused because of lack of procedures.

Quality as a category appeared vague to the manager and he felt it was best if better categories were established.

The interviewer mentioned the card sorting study associated with aircraft maintenance categories that were carried out with auditors and quality assurance representatives. The manager asked the question as to why so much weight was on what the auditors have to say about organizational categories. He mentioned that the auditors are employed to conduct audits and the quality assurance representatives are employed to conduct surveillance on vendor locations during aircraft maintenance processes. Something as complicated as organizational categories is the job of the managers and the higher management to conceptualize and finalize. The manager was not surprised about the inadequacy of the second set of categories (low cost, medium cost, high cost, and no impact).

The interviewer inquired about the potential categories the technical audits manager would want the design team to incorporate. The manager said the implications at an organizational level would include factors such as Safety and Economic. Economic would be at four levels: low, medium, high, and no impact. The manager expressed the need to have a dollar value associated with the levels of the category of Economic, but also informed the interviewer that this was difficult since the dollar value associated with findings are subjective. The manager felt that the four levels of low, medium, high, and no impact were a good start. The manager mentioned that Economic is a very general category because every maintenance finding is either associated with cost or injury, and thus ultimately it is either related to Economical or Safety. He explained further by stating that: “You either crash an airplane or go out of business. It is very difficult to assign a dollar value for a finding.”

At this time, the interviewer sensed a similarity between the requirements of the manager and the risk management matrix recommended by the manager of the internal audits department. The technical audits department manager was shown the risk management matrix. The manager was impressed and commented, “This is something which we could also use.”

The internal audit RMM seemed a good starting point to the manager. He said that the technical audit department might make changes to this matrix as time goes by, depending on peculiar situations specific to the technical audits department.

The manager reiterated the point that they are responsible to set standards to decide what to measure, how to measure, and then analyze the associated data.

Interview with the Surveillance manager

The interviewer reminded the manager of categories such as Human Factors, Quality, and Economic.

The manager was no longer confident in these categories. The manager derived an analogy about what ISO standards help achieve, which is quality and oversight of processes. The common denominator according to him was ‘risk’. He mentioned how everything in the industry is risk based. The manager mentioned that if risk is determined and evaluated properly, then risk indicates what needs to be done.

The manager mentioned that he wanted to remove subjectivity from the minds of quality assurance representatives regarding the risk level associated with a surveillance finding. This immediately made him mistrust the categories such as low cost, medium cost, high cost, and no impact.

The manager wanted choices to be hard coded for the quality assurance representatives. Yet the manager expressed the need to have the flexibility to adjust the risk level choices of the quality assurance representatives based on his interpretation. The manager stressed the fact that he mostly struggles with the subjectivity of the interpretations of the quality assurance representatives.

The manager mentioned that he did not have a starting point for risk levels to indicate the impact of a surveillance finding at an organizational level.

Based on the interviewer’s understanding of the problem at hand, the internal audit RMM was recommended to the manager who agreed and said that it was a good starting point. The manager was informed about the acceptance of the RMM as a solution by the manager of the technical audits department. The manager was hopeful that a common approach to resolve the issue would allow the managers to understand problems across the organization, and not be focused on their department alone.

Resolution

Based on information provided on the RMM by the Internal Audits Manager, and the subsequent agreement of managers of the Technical Audits and Surveillance Departments, it was decided that the Risk Management Matrix would be used to classify surveillance and audit findings in terms of implication of a maintenance finding at an organizational level. The RMM allows internal auditors and their manager to classify an audit finding in terms of ‘risk’, which is associated with a severity of the finding and a probability associated with the occurrence of the finding

Discussion

This research transformed qualitative maintenance data to establish performance measures at two levels. The research utilized the expertise of stakeholders at each stage of the research. The research did not assume any structure of the qualitative maintenance data.

Multidimensional Scaling was utilized to generate performance measures for aircraft level impact. Linear regression was then utilized to interpret the multidimensional spatial solution and validate the dimensions. These dimensions are the performance measures that indicate the impact of maintenance and audit findings.

A card sorting study was initially conducted to uncover the potential performance measures for Organizational Categories, indicating impact of maintenance and audit findings on organizational cost. There was a flaw with this study. The card sorting study was conducted with quality assurance representatives from the surveillance department, technical auditors, and internal auditors. The quality assurance representatives and auditors admitted that they do not think about maintenance findings in terms of organizational cost. The researcher then involved quality assurance managers to uncover a RMM, which would be used to interpret maintenance findings at an organizational level in terms of ‘Risk’. In the RMM, ‘Risk’ is a function of probability of occurrence and severity of the finding.

The research involved quality assurance maintenance data from our industry partner for this research. It is important to validate the identified performance measures with quality assurance personnel from other organizations.

References

Adams, S.T., & Kirwan, B. (1997). Human reliability data requirements. *Disaster Prevention and Management, 6*(5), 318-335.

Beyer, H., & Holtzblatt, K. (1997). *Contextual design: Defining customer-centered systems*. San Francisco: Morgan Kaufmann.

Blomberg, J.L., & Henderson, A. (1990). Reflections on participatory design: Lessons from the Trillium Experience. *Conference on Human Factors in Computing Systems: Proceedings of the SIGCHI conference on Human Factors in Computing Systems: Empowering People,* Seattle, Washington, U.S.A., 353-360.

Brence, J.R., & Brown, D.E. (2002). Data mining corrosion from Eddy Current Non-Destructive Tests. *Computers and Industrial Engineering, 43*(4), 821-840.

Cho, B.R., Kim, Y.J., Kimbler, D.L., & Phillip, M.D. (2000). An integrated joint optimization procedure for robust design and tolerance design. *International Journal of Production Research*, *38*(10), 2309-2325.

Green, S.B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*, 449-510.

Gronbaek, K., Kyng, M, & Mogensen, P. (1997). Toward a cooperative experimental system development approach. In Kyng, M., and Mathiassen (eds.): *Computers and Design in Context,* Cambridge, MA: MIT Press, 201-238.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco: Kaufmann.

Holmgren, M. (2005). Maintenance-related losses at the Swedish rail. *Journal of Quality in Maintenance Engineering, 11*(1), 5-18.

Jones, L.E., & Young, F.W. (1972). Structure of a social environment: Longitudinal individual differences scaling of an intact group. *Journal of Personality and Social Psychology, 24*, 108-121.

Kensing, F., & Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work, 7*(3-4), 167-185.

Kim, Y.J., & Cho, B.R. (2000a). Economic considerations on parameter design. *Quality and Reliability Engineering International*, *16*(6), 501-514.

Kim, Y.J., & Cho, B.R. (2000b). Economic integration of design optimization. *Quality Engineering*, *12*(4), 561-567.

Kruskal, J.B., & Wish, M. (1983). *Multidimensional scaling*. Newbury Park: Sage Publications.

Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259-284.

Luck, R. (2003). Dialogue in participatory design. *Design Studies, 24*(6), 523-535.

Muller, M.J. (1992). Retrospective on a year of participatory design using the PICTIVE technique. *Conference on Human Factors in Computing Systems: Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* Monterey, CA, U.S.A., 455-462.

Muller, M.J., Wildman, D.M., & White, E.A. (1993). Taxonomy of participatory design practices: A brief practitioner's guide. *Communications of the ACM, 36*(4), 24-28.

Patankar, M.S., & Taylor, J.C. (2003). Posterior probabilities of causal factors leading to unairworthy dispatch after maintenance. *Journal of Quality in Maintenance Engineering, 9*(1), 38-47.

Rosenberg, S., Nelson, C., & Vivekananthan, P.S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*(4), 283-294.

Shyur, H., Luxhoj, J.T., & Williams, T.P. (1996). Using neural networks to predict component inspection requirements for aging aircraft. *Computers and Industrial Engineering, 30*(2), 257-267.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights: Allyn & Bacon.

Taguchi, G. (1986). *Introduction to quality engineering*. New York: Krauss International Publication.

Wang, C., & Hwang, S. (2004). A Stochastic maintenance management model with recovery factor. *Journal of Quality in Maintenance Engineering, 10*(2), 154-165.

Weckman, G.R., Shell, R.L., & Marvel, J.H. (2001). Modeling the reliability of repairable systems in the aviation industry. *Computers and Industrial Engineering, 40*(1-2), 51-63.

Yoon, W.C., & Kim, Y.S. (1996). Aiding the analysis of human actions in large-scale systems: An intelligent interface approach. *Computers and Industrial Engineering, 30*(3), 569-577.

International Journal of Applied Aviation Studies, Volume 7, Number 1

Runway Incursions: A Critical Examination of Airport Driver Training Methods

William B. Rankin, II
Florida Memorial University
15800 NW 42 Avenue
Miami Gardens, FL 33054
954-806-9840

Abstract

According to the FAA Runway Safety Report (2004), vehicle deviations accounted for 20% (291 events) of all runway incursions during the period of 2000 through 2003. The focus of this quantitative correlation study examined if a relationship existed between the methods used for airport movement area driver training and the number of incursions at the Operational Evolution Plan (OEP-35) U.S. towered airports. Airport driver training officials at the OEP-35 airports were surveyed using a five-point Likert-type survey. The data from this study suggested a relationship exists between the methods used for airport driver training and the number of runway incursions at the largest U.S. towered airports. The data from this study might help to reduce airport liability exposure, property damage, and lower airport liability insurance cost to U.S. airport owners.

Introduction

Since the mid-1920s, commercial aviation in the United States has achieved a remarkable safety record. Within the National Airspace System (NAS), thousands of passenger trips and aircraft operations are completed safely every year (Federal Aviation Administration, 2002a). The growing pressure for increased operational rates to reduce system delays, combined with the complexity of airport operations and the requirement for precise timing, combine to make the airport movement area surfaces unforgiving of errors by pilots, air traffic controllers, and vehicle drivers (FAA, 2002b).

According to Clarke (2002), the Federal Aviation Administration (FAA) has developed several training programs for pilots and air traffic controllers to make each group more aware of runway incursion problems. In addition, the FAA instituted Standardized Taxi Routes (STRs) by FAA Order 7110.116, to assist pilots and air traffic controllers with surface movement of aircraft. Finally, air traffic controllers are required to maintain a high level of runway incursion awareness

---

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

Return to Table of Contents

through a monthly computer-based recurrent training program titled *Preventing Runway Incursions.*

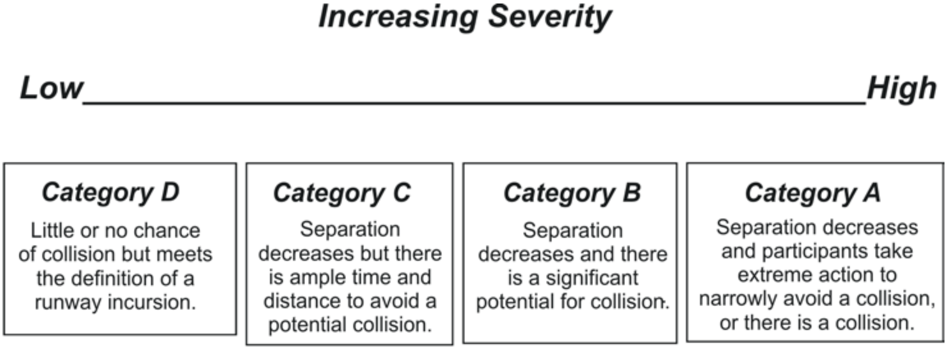
Rankin (1994) identified training of ground vehicle operators as the most effective FAA initiative to reduce runway incursions; however, ground vehicle operator training is conspicuously absent from mention in most literature, even though vehicle operators traverse airport movement areas on a daily basis. An earlier study by Hacker (2002) supported this position. Hacker stated:

> At present, no real standards exist to train individuals for preventing runway incursions. Pilot deviations, operational errors, and vehicle deviations increased 34.3% from calendar year 1999 to 2000 and yet no sections of Federal Aviation Regulation Part 139 or any 150 series advisory circulars even address runway incursions or preventative methods to incorporate into a training program. (p. 31)

On June 21, 2002, FAA issued Advisory Circular (AC) 150/5210-20 to provide guidance to airport operators in developing training programs for vehicle ground operations. This was the first advisory circular providing airport operators with a list of training topics to include in a ground vehicle operator-training curriculum (FAA, 2002a).

The FAA (2004b) defined runway incursions as “Any occurrence at an airport involving an aircraft, vehicle, person, or object on the ground that creates a collision hazard or results in a loss of required separation with an aircraft taking off, landing, or intending to land” (p. 9). The NAS continues to experience approximately one runway incursion per week, which is classified as significant or a barely avoided collision (FAA 2004b).

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



Category D	Category C	Category B	Category A
Little or no chance of collision but meets the definition of a runway incursion.	Separation decreases but there is ample time and distance to avoid a potential collision.	Separation decreases and there is a significant potential for collision.	Separation decreases and participants take extreme action to narrowly avoid a collision, or there is a collision.

Figure1. Runway incursion categories by increasing severity. (FAA, 2004)

Statement of Problem

The study addressed the problem of runway incursions at the largest U.S. towered airports – the FAA designated Operational Evolution Plan (OEP-35) airports. The focus of this quantitative correlation study was to examine if a relationship exists between the methods used for airport movement area driver training and the number of runway incursions for any class of runway incursions (as determined by the FAA) at the largest U.S. towered airports.

According to the *FAA Runway Safety Report* (2004), vehicle deviations accounted for 20% (291 events) of all runway incursions during the four and a half-year period of 2000 through 2003. Of the 189 category A and B runway incursions during this period, 16% or 30 events were vehicle deviations. Although there has been a slight decrease in the overall number of vehicle deviations nationally, the number of category A and B vehicle deviations increased slightly from 2001. In 2001, there were 83 vehicle deviations, with five of the events being classified as category A and B incursions. In 2003, the number of vehicle deviations decreased to 60 events; however, nine of the events were classified as category A and B incursions (FAA, 2004).

Although vehicle deviations represent a smaller portion of the total U.S. runway incursions, the potential risk in the terms of loss of life is significant. The most serious runway incursion to date (a pilot deviation) occurred in Tenerife, Canary Island, on March 27, 1977, killing 583 people, and ranking as the worst disaster in aviation history (Clarke, 2002).

This study is significant in that no previous study has been conducted on the relationships between the methods used for airport movement area driver training and the number of runway incursions. Leaders at the FAA have determined that the 35 busiest U.S. towered airports have twice the average number of reported runway incursions than the rest of the U.S. towered airports (FAA, 2004). The reduction of just one runway incursion could reduce the potential of a catastrophic event.

Purpose of Study

The purpose of this study was to identify the more effective method of airport movement area driver training that will reduce the potential for runway incursions at the largest U. S. towered airports. There are currently two methods of airport movement area driver training being used at the OEP-35 airports. These include traditional classroom movement area driver training (traditional) and the American Association of Airport Executives (AAAE) interactive computer-based airport movement area driver training. The AAAE has a patent on the interactive computer based method of delivery.

Definition of Terms

Air traffic control:

A service operated by an appropriate authority to promote the safe, orderly, and expeditious flow of air traffic (FAA, 2002a).

Loss of separation:

This is an occurrence or operation that results in less than Federal Aviation Regulation (FAR) prescribed separation between aircraft, vehicles, or objects (Clarke, 2002).

Movement areas:

The runways, taxiways, and other areas of an airport that are used for taxiing, takeoff, and landing of aircraft, exclusive of loading ramps and aircraft parking areas (FAA, 2002a).

Non-movement areas:

Taxiways, aprons, and other areas not under the control of air traffic or at airports without an operating airport traffic control tower (FAA, 2002a).

Operational deviation (error):

This is an action of an air traffic controller that results in a loss of separation, i.e., less than the FAR-required minimum separation between an aircraft and obstacles (vehicles, equipment, personnel) on runways (Clarke, 2002).

Pilot deviation:

An action (or inaction) of a pilot that violates any FAR. The most prevalent examples of pilot deviations occur when a pilot fails to obey air traffic control (ATC) instructions to hold short of an active runway or when a pilot selects and uses a specific taxiway without proper clearance (Clarke, 2002).

Vehicle deviation:

This is the entry or movement on the runway or taxiway area by a vehicle that has not been authorized by ATC. As a side note, included is the taxiing aircraft operated by non-pilots. For the purpose of this study, a vehicle deviation is the same as a vehicle/pedestrian deviation (Clarke, 2002).

Review of Literature

The review of literature examines training effectiveness from several aspects. A review of literature on the *Runway Safety Blueprint 2002-2004* addressed the primary causes for runway incursions and the complexities involved in solving runway incursions. The goals of the FAA *Runway Safety Blueprint* (2002a) are consistent with those identified by Rankin (1994) for all airports and included the following:

1. Develop and distribute runway safety education and training materials to controllers, pilots, and all other airport users.
2. Increase surface safety awareness throughout the aviation community.
3. Assess and modify procedures to enhance runway safety.
4. Improve runway safety data collection, analysis, and dissemination.
5. Identify and implement enhancements to improve surface communications.
6. Increase situational awareness on the airport surface.
7. Support and deploy new technologies that reduce the potential for collision.
8. Implement site-specific runway safety solutions in coordination with local aviation communities. (p. 4)

A review of literature on distance education and computer-based interactive training addresses knowledge gained by other researchers on traditional versus computer-based training, and the skills transfer capabilities of the various methods of training.

Since training has been historically linked to education, it is not surprising to see computer-based interactive training linked to distance education. Nanney (n.d.) stated distance learning can be interactive or non-interactive:

> Interactive learning can be synchronic or asynchronic, or a combination of the two. In synchronic learning the teacher and student perform interactively at the same time on the same subject and in every learning action, they perform as in the traditional classroom. Non-interactive learning is mainly represented by the World Wide Web where the media transfers the knowledge to the learner. Distance education in its many forms is organized so that it is at some point between totally interactive and completely non-interactive learning. (Nanny, n.d., p. 1)

According to Filipczak (1996), computer-based interactive training can be the most effective way to train. Almost every example of effective computer-based training is based on multimedia simulation or a simulated environment. Filipczak stated:

> The most powerful learning environments are simulators, but only when learning is designed into the environment. The point of multimedia training is not to create an interactive environment for its own sake, but to create a dynamic educational construct where people can learn. (p. 1)

In addition, according to Rohland (1996), “Computerization is the trend in training. More companies are seeing that taking people into the classroom is not the answer” (p. 1).

Kirkpatrick (1983) addressed the four *aspects of training* evaluation in Kirkpatrick’s model with respect to training effectiveness (p. 44). According to Tidler (1999), Kirkpatrick’s model is widely accepted by the American Association for Training Development, and there are four *aspects of training* in Kirkpatrick’s model with respect to training effectiveness (p. 44). These aspects include:

1. Reactions--What trainees’ say about the value of the training.
2. Learning--Objectives met, knowledge, and skills learned.
3. Behavior--The skills acquired are implemented on-the-job.
4. Results--Impacts on job performance. (Kirkpatrick, 1983, ¶ 6)

Kirkpatrick’s model was selected as the model most appropriate to use for the development of a model for the study of airport driver training methods at the largest U.S. towered airports.

Highlights of Methodology

This study focused only on the runway incursion problem caused by vehicle deviations, which is under the purview of airport operators. This research was intended to be primarily a descriptive and correlational (non-experimental) analysis of the relationships, if any, that exist between the methods used for airport movement area driver training and the number of runway incursions at the largest U.S. towered airports.

For the purposes of this study, a quantitative and limited qualitative methodology was used. A five point Likert-type survey instrument was used to collect the necessary data and qualitative responses to five end-of-survey questions (see Appendix A). The statistical analyses used in the study included multivariate analysis of variance (MANOVA).

Limitations of Study

According to Wells and Rodriquez (2004), the National Airspace System Plan (NASP or NPIAS) contains a listing of more than 3,000 public funded airports in the United States. Of the 3,000 plus public funded airports listed in the NASP, only 490 airports have operating control towers, and only towered airports report runway incursions (FAA, 2004). With respect to the towered airports, FAA (2004) stated commercial aviation aircraft operations at the OEP-35 airports are predominantly commercial aircraft, and account for "the majority (87 percent) of category A and category B runway incursions" (p. 36).

This study was limited to an analysis of vehicle deviations that cause runway incursions at the OEP-35 airports. According to the FAA (2004), "vehicle/pedestrian deviations represented 18 percent of the runway incursions at the OEP-35 airports, which is in proportion to their national representation (20 percent)" (p. 36).

There were several potential limitations with respect to the survey instrument. These included (a) the effective sample size of participants, (b) the accuracy of the data provided by the participants, and (c) the pitfall of correlation versus causation for forming conclusions.

Statement of Hypotheses

This study focused on the relationship between the methods used for airport driver training and the number of runway incursions at the largest U.S. towered airports. In concert with the stated research question: Does a relationship exist between the methods used for airport driver training and the number of runway incursions for any class of incursions at the largest U.S. towered airports, there was one null and alternative hypotheses framed for this study. The hypotheses were formulated as follows:

1. HO_1: There is no relationship between the methods used for airport movement area driver training and the number of runway incursions for any class of incursions.
2. H_{11}: There is a relationship between the methods used for airport movement area driver training and the number of runway incursions for any class of incursions.

Description of Materials and Instruments

Past studies using Kirkpatrick's model (Bledsoe, 1999; Kirkpatrick, 1983; Naugle, Naugle, & Naugle, 2000; Tidler, 1999) focused only on the issues of face and content validity, and not on reliability. These studies relied primarily upon the review of the literature and input from training professionals. Both face validity and content validity were achieved in that the experts involved in the refinement of the tools used to measure training effectiveness concurred that the instruments effectively measured what they set out to measure. In addition, the questions in these studies were carefully scrutinized to determine whether they contributed to the measurement that the instruments were designed to quantify. Factor analysis was conducted to assess construct validity.

When evaluating an attribute of a trainee such as driver training ability, a measurement device is needed. This is generally a scale or test that consists of multiple individual items. The response to each item is graded and summed, resulting in scores for each case. For this study, a pilot survey questionnaire was used to gather the preliminary data for measurement by the appropriate statistical analyses. The preliminary data was then used to test the scale prior to initiation of the final data collection.

For the purposes of this study, a reliability analysis that allowed the researcher to study the properties of the measurement scale and the items that make up the scale was used. The reliability analysis procedure calculated a number of commonly used measures of scale reliability and provided information about the relationships between individual items in the scale. Reliability was evaluated using Cronbach's Alpha (Norusis, 2003).

Selection of Participants

The population for the study was comprised of employees that completed airport movement area driver training at the 18 airports responding to the survey in the FAA Operational Evolution Plan (OEP-35). Population data included archival data from the FAA on vehicle runway incursions (categories A through D), by year, for each of the dependant variables, and survey data on both the independent and intermediary independent variables from targeted participants at the OEP-35 airports. Targeted participants included 390 randomly selected employees who have successfully completed airport movement areas driver training and who are authorized to drive vehicles onto and within the airport movement areas.

While runway incursion data was collected from the FAA database, a scaled survey instrument was used to gather the data on the independent and intermediary independent variables (questions 1-19) associated with the four aspects of training effectiveness. Literature obtained from the AAAE identified the training methods in effect by airport over the study period (AAAE, 2005b).

Discussion of Data Processing

Power analysis software obtained from the UCLA Department of Statistics was used to estimate the required number of *completed* surveys. The calculations showed that at least 194 completed surveys needed to be collected from participants at the OEP-35 airports to estimate the mean response values for questions 1 through 19 within a desired precision of .10 (University of California, 2005).

FAA (2004) considers runway incursions rare events relative to total aircraft flights over finite periods of time (5.6 incursions per half million aircraft flights per year). According to Aczel and Sounderpandian (2002) "if we count the number of times a rare event occurs during a fixed interval, then that number would follow a Poisson distribution" (p. 151). Using software obtained from the UCLA Department of Statistics web page (University of California, 2005), a Poisson power analysis was used to estimate the number of years of runway incursion data needed from the 2004 FAA *Runway Safety Report*.

Estimated marginal means from SPSS© output were used to evaluate statistical significance differences among the means of the dependent variables – runway incursion categories A through D.

The statistical analysis used in the study included multivariate analysis of variance (MANOVA). The significance of independent variables associated with aspects 1 through 4 as well as the effects of individual intermediary independent variables associated with the aspects were tested using the multivariate procedures. In addition, the effects of covariates and covariate interactions with factors were included. For MANOVA, the intermediary independent variables and demographic variables were specified as covariates (Norusis, 2003).

Since more than one dependent variable was specified, the multivariate analysis of variance using Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root criterion with approximate *F* statistic was provided as well as any subsequently needed univariate analysis of variance for each dependent variable. Estimated marginal means gave estimates of mean values for the dependent variables in the model (Norusis, 2003).

Table 1 depicts the method of analysis that was used to analyze the research question and sets of variables. For purposes of the study, traditional airport movement area driver training was identified as (TRADmethod) and AAAE interactive computer-based airport movement area driver training was identified as (AAAEmethod). The categories of runway incursions (A-D) were identified as (INCURa), (INCURb), (INCURc), and (INCURd).

Table 1
Research Question, Study Variables, and Methods of Analysis

Research Question	Study Variables	Level of Measurement	Methods of Statistical Analysis
Does a relationship exist between the methods used for airport movement area driver training and the number of runway incursions for any class of runway incursions at the largest U.S. towered airports?	TRADmethod AAAEmethod INCURa INCURb INCURc INCURd	Nominal Nominal Ratio Ratio Ratio Ratio	MANOVA

The model for the relationships between the independent variable, aspects (intermediary independent variables), research questions, and dependent variables is illustrated in Figure 2. The model shows that the independent variable consisted of two airport movement area driver training methods – AAAE interactive computer-based airport movement area driver training, and traditional airport movement area driver training. There were four aspects associated with the training method. These included (a) aspect 1- learning objectives met, (b) aspect 2 – knowledge increase, (c) aspect 3 – on-the-job confidence, and (d) aspect 4 - effectiveness of materials and methods. Questions 1-6 of the survey instrument addressed aspect 1; questions 7-8 addressed aspect 2; questions 9-14 addressed aspect 3; and questions 15-19 addressed aspect 4. Aspects 1 through 4 are the intermediary independent variables, which affect the dependent variables – the four categories of runway incursions A through D.

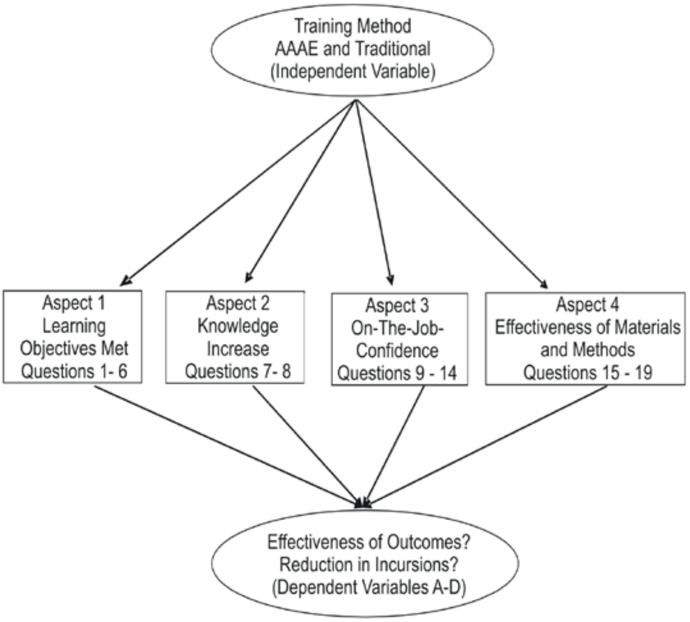


Figure 2. Study model depicting the relationship of the variables and outcomes. (Tidler, 1999)

Findings

Reliability of the Survey Instrument

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

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

In the case of this study, SPSS© software was used to calculate the Cronbach's Alpha value of 0.864 shown in Table 2 for the 19 survey questions used to study driver training methods.

Table 2

Cronbach's Alpha Reliability Statistics

Cronbach's Alpha	Number of Items
.864	19

Distribution of the Dependant Variables

The second step was to determine what distribution the dependent variables (runway incursion categories A through D) followed. As previously stated, FAA (2004) considers runway incursions rare events relative to total aircraft flights over finite periods of time (5.6 incursions per half million aircraft flights per year). According to Aczel and Sounderpandian (2002) "If we count the number of times a rare event occurs during a fixed interval, then that number would follow a Poisson distribution" (p. 151).

Relationships of the Methods used for Airport Driver Training

The third step was to examine the nature of the relationships that existed between the number of runway incursions and the two methods used for airport driver training. The relationships were analyzed using multivariate analysis of variance (MANOVA).

Multivariate Analysis of Variance (MANOVA) Results

MANOVA Analysis of the Effects of Training Methods

MANOVA identified the effect of the independent variable, *method of training*, across all classes of runway incursions, as statistically significant at the 0.000 level for airport movement area driver training methods.

MANOVA Analysis of the Effects of the Intermediary Independent Variables

The MANOVA analysis also identified the effects of the intermediary independent variables in the model. The variable, *understanding of ATC language, clearances, instruction, and light signals*, was found to be statistically significant at the 0.046 level. This variable measured the participants' understanding of air traffic control (ATC) language, and interpretation of clearances, instructions, and light signals issued by ATC. The variable, *knowledge and skill on movement area driving before completing driver training*, was found to be statistically significant at the 0.001 level. This variable measured the participants' knowledge of the skills needed to drive on the movements areas safely before completing driver training. The variable, *knowledge and skill on movement area driving after completing driver training*, was found to be statistically significant at the 0.015 level. This variable measured the participants' knowledge of the skills needed to drive on the movements areas safely after completing driver training. The variable, *understand notice to airmen*, was found to be statistically significant at the 0.003 level. This variable measured the participants' on-the-job-confidence to interpret notices published that advise of movement area hazards. The variable, *materials used provided enough information,* was found to be statistically significant at the 0.000 level. This variable measured the participants' belief that the materials used in training were adequate to train drivers to operate vehicles safely on movement areas. The variable, *overheads used are clear and easy to follow*, was found to be statistically significant at the 0.000 level. This variable measured the participants' belief that the overhead slides used in training were adequate to train drivers to operate vehicles safely on movement areas. Finally, the variable, *driver training video used*, was found to be statistically significant at the 0.008 level. This variable measured the participants' belief that the video(s) used in training were adequate to train drivers to operate vehicles safely on movement areas. See Appendix B for the multivariate (MANOVA) results.

Pair wise Comparisons Test

Table 3 provides the SPSS© results summary of the pair wise comparisons for the two training methods for each of the four incursion categories.

Table 3
Pair wise Comparisons

Dependent Variable	**(I) Method Used**	**(J) Method Used**	**Mean Difference (I-J)**	**Std. Error**	**Sig.(a)**	**95% Confidence Interval for Difference**	
Category						**Lower Bound**	**Upper Bound**
A	Traditional	AAAE Interactive	.399(*)	.080	.000	.241	.556
B	Traditional	AAAE Interactive	1.452(*)	.284	.000	.893	2.011
C	Traditional	AAAE Interactive	1.792(*)	.453	.000	.898	2.685
D	Traditional	AAAE Interactive	.529	.651	.418	-.755	1.813

NOTE: Based on estimated marginal means. * The mean difference is significant at the .05 level.

Estimated Marginal Means

As shown in Table 4, MANOVA identified the estimated marginal means (estimates of mean values for the dependent variables) for each training method by runway incursion categories A through D. For the AAAE interactive computer-based airport movement area driver training method, the estimated marginal mean for the number of category A runway incursions was 0.129; the estimated marginal mean for the number of category B runway incursions was 0.809; the estimated marginal mean for the number of category C runway incursions was 2.975; and the estimated marginal mean for the number of category D runway incursions was 4.323. For the traditional airport movement area driver training method, the estimated marginal means for the number of category A runway incursions was 0.528; the estimated marginal mean for the number of category B runway incursions was 0.809; the estimated marginal mean for the number of category C runway incursions was 4.767; and the estimated marginal mean for the number of category D runway incursions was 4.852. As shown in Appendix C, profile plots of the estimated marginal means allowed for visualization of the training methods relationships.

Table 4
Estimated Marginal Means

Dependent Variable	Method Used	Mean	Std. Error	95% Confidence Interval	
Category				Lower Bound	Upper Bound
A	Traditional	.528	.039	.451	.605
	AAAE Interactive	.129	.065	.000	.258
B	Traditional	2.260	.139	1.986	2.535
	AAAE Interactive	.809	.232	.352	1.265
C	Traditional	4.767	.223	4.328	5.206
	AAAE Interactive	2.975	.370	2.245	3.705
D	Traditional	4.852	.320	4.221	5.483
	AAAE Interactive	4.323	.532	3.274	5.372

Analysis and Evaluation of Findings

Reliability of the Survey Instrument and Distribution of Dependent Variables

The Cronbach's Alpha value of 0.864 implied a high degree of reliability among the 19 questions used in the survey instrument. Runway incursions are rare events relative to total aircraft operations over finite periods of time (5.6 incursions per half million aircraft operations per year). As a result, the dependant variables (runway incursion categories A through D) follow a Poisson distribution.

Multivariate Analysis of Variance (MANOVA)

MANOVA identified the independent variable, *method of training*, as statistically significant at the 0.000 level. Pair wise comparisons in MANOVA identified traditional airport movement area drive training as statistically different from AAAE interactive computer-based airport movement area driver training at a statistically significant level of 0.000 for runway incursion categories A through C. The only exception, in the pair wise comparisons test, was category D runway incursions, which were not statistically significant at the 0.418 level. (see Table 3) Finally, estimated marginal means values for runway incursion categories A through D supported the finding that those airports using AAAE interactive computer-based airport movement area driver training have the propensity for fewer runway incursions caused by vehicle deviations for all categories of incursions (see Table 4).

As a result, MANOVA analyses supported the alternative hypothesis that there is a relationship between the two methods used for airport movement area driver training and the number of runway incursions for any class of incursions, with the exception of category D incursions.

Summary

A Cronbach's Alpha value of 0.864 implied a high degree of reliability for the 19 questions used in the survey instrument. Since runway incursions are rare events, it was determined that the dependant variables studied (runway incursion categories A through D) follow a Poisson distribution.

Multivariate procedures were used to analyze the research question: Does a significant relationship exist between the methods of airport driver training and the number of runway incursions for any class of incursions at the largest U.S. towered airports? Since more than one dependent variable was specified, the MANOVA using Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root criterion with approximate *F* statistic was provided. These four statistics were found to support the significance of the training method used. Estimated marginal means identified the estimates of mean values for the dependent variables in the model (see Table 4).

Overall, the quantitative analysis supported the alternative hypothesis that there is a relationship between the methods used for airport driver training and the number of runway incursions for most classes of incursions.

Conclusions

The MANOVA analysis supported the alternative hypothesis that there is a relationship between the methods used for airport movement, area driver training, and the number of runway incursions for all incursion classes, except class D. Other than publication of FAA Advisory Circular 150/5210-20 on airport vehicle surface operations, airport movement area driver training method is not specifically addressed in the 40 runway incursion prevention initiatives outlined in the *FAA Runway Safety Blueprint 2002-2004* (FAA, 2004). Efforts by the FAA to date have focused primarily on air traffic controllers and airline pilots, although 20% of the annual accidents at the OEP-35 airports involve airport vehicles deviations.

Most of the OEP-35 airports use traditional airport driver training. The data from this study suggested that for those airports using traditional airport movement area driver training, the propensity for the number of runway incursion accidents, categories A through C, is more likely than at those airports that are using AAAE computer-based airport movement area driver training. This implies that the propensity for runway incursions due to vehicle deviations is higher overall in the U.S. than would be the case if the AAAE interactive computer-based airport movement area driver training method was implemented at all the OEP-35 airports. With the exception of category D runway incursions, these finding are inconsistent with the generally accepted thinking within the airport industry that both methods of airport movement area driver training are equally as likely to reduce runway incursions.

Qualitative responses indicated that English is a second language for many employees with airport movement area driving privileges. Since English is the industry adopted language for aviation operations worldwide, this finding is problematic. This is evidenced by Clarke (2002) who stated, "The use of consistent terminology (in ATC communication) is recommended for all involved" (p. 14).

Industry Implications

This study identified AAAE interactive computer-based airport movement area driver training as the more effective method of airport movement area driver training that will reduce the potential for runway incursions at the largest U. S. towered airports. Since the data suggested that there is potential to reduce runway incursions by replacing one driver training method with another, the potential exist to reduce airport liability exposure at all U. S. airports. Other benefits would include a reduction in property damage, and an overall lowering of airport liability insurance cost to airport owners.

Runway incursion data from the *FAA Runway Safety Report* (2004) indicated that for the 18 airports that participated in this study, those airports reported 216 runway incursions over the study period of 2000 through 2003 -- an average of more than four runway incursions per airport per year (FAA, 2004). Each runway incursion has the potential for loss of life and property damage.

Recommendations

Accordingly, it is recommended that the Federal Aviation Administration should mandate that all the OEP-35 airports acquire and implement the AAAE interactive computer-based airport movement area driver training system over the next two-to-three year timeframe, or as quickly as the systems can be acquired and installed. This initiative should be added to the *FAA Runway Safety Blueprint 2002-2004* and implemented through an amendment of Federal Aviation Regulation Part 139. Subsequent to the later action, airport operators should be required by an amendment to Federal Aviation Regulation Part 139 to address avoidance of vehicle deviations in their Airport Certification Manual. Concurrent with these actions, FAA should designate the Office of Runway Safety within the FAA as the focal point for this effort. Additionally, the Office of Safety Services within the Air Traffic Organization (ATO) should play a key role in this effort.

Further study of the airport movement area driver training is needed to understand why AAAE interactive computer-based airport movement area driver training is not more effective with respect to category D runway incursions and what changes, if any, could be made in the AAAE training curriculum or system to make it more effective. One explanation for the lack of training effectiveness may be similar in nature to motorist failing to observe proper and lawful automobile traffic controls such as speed limits, stop signs, and traffic lights, etc. Although the vast majority of the driving public operates in a safe and proper manner, a small percentage fails (by choice, ignorance, or inattention) to observe the rules of the road.

Study, education, and strict enforcement are the tools currently being used by airport operators to address the problem of vehicle deviations. This system of addressing vehicle deviations is sometimes called *study, educates, enforces* (SEE) and has been successful in many areas, not just aviation (Clark, 2002).

A study of airport movement area driver training effectiveness by geographic regions may also provide additional insights into problem of runway incursions by specific regions of the United States. For example, certain regions (like South Florida) may have racial and ethnic differences, which lead to communications barriers that are not experienced in other regions of the United States. This may mean that educational materials need to be translated into other languages in order for the materials to be delivered effectively, or that English competency test should be required by FAA regulation before any employee may seek airport movement area driving privileges.

An additional recommendation for subsequent research is that the scope of a similar study should be expanded to all U.S. towered airports, including those outside the scope of this study. Finally, the current runway safety initiatives contained in the *FAA Runway Safety Blueprint 2002-2004,* should be evaluated and ranked in the order of there effectiveness by a survey of industry officials.

References

Aczel, A. D., & Sounderpandian, J. (2002). *Business statistics* (5th ed.). New York: McGraw-Hill Irwin.

American Association of Airport Executives (AAAE). (2005b, February 17). *Interactive employee training (IET).* [Brochure]. Retrieved March 26, 2005, from http://www.aaae.org

Bledsoe, M. D. (1999). Correlations in Kirkpatrick's training evaluation model. *Dissertation Abstracts International*, 60/01, 54. (UMI No. 9937553)

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

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

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

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

Filipczak, B. (1996, November). Endangered! The nature of computer interactivity. *Training, 33*(7), 52-58. Retrieved March 30, 2005, from ProQuest database.

Hacker, C. C. (2002). *Preventing runway incursions through education and training: An in-depth analysis.* Unpublished AAAE accreditation thesis, American Association of Airport Executives. Alexandria, Virginia.

Kirkpatrick, D. L. (1983, November). Four steps to measuring training effectiveness. *The Personnel Administrator*, *28*(11), 19-24. Retrieved April 1, 2005, from ProQuest database.

Nanney, B. A. (n.d.). Distance education. Retrieved March 27, 2005, from http://www2.gsu.edu

Naugle, K. A., Naugle, L. B., & Naugle, R. J. (2000, Fall). Kirkpatrick's evaluation model as a means of evaluating teacher performance. Retrieved April 3, 2005, from FirstSearch database.

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

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

Rohland, P. (1996, April 1). New techniques spice up corporate training. *Eastern Pennsylvania Business Journal 7*(13), 6-19. Retrieved March 30, 2005, from ProQuest database.

Tidler, K. L. (1999). Evaluation of continuing medical education using Kirkpatrick's four levels of evaluation. *Dissertation Abstracts International*, 61/01, 59. (UMI No. 9958848)

University of California, Los Angeles (2005). *UCLA calculators.* Retrieved January 26, 2006, from http://calculators.stat.ucla.edu

Wells, A. T., & Rodriques, C. C. (2004). *Commercial aviation safety* (4th ed.). New York: McGraw-Hill.

Appendix A

Driver Training Survey

Driver Training Effectiveness Survey

Name of Airport______________________Date_______

Guidelines: Read each of the following questions and circle the number that most appropriately represents your opinion. For the qualitative questions on the final page, please clearly print your responses in the space provided. N/A is not applicable.

LEARNING OBJECTIVES MET	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	N/A
When you completed the Driver training, you were able to:						
1. Understand the meaning of runway/taxiway signs, surface markings, and lighting	1	2	3	4	5	N/A
2. Understand ATC language, clearances, instructions, and light signals	1	2	3	4	5	N/A
3. Understand Notice to Airmen (NOTAMs)	1	2	3	4	5	N/A
4. Understand and use airport diagrams to navigate on the surface	1	2	3	4	5	N/A
5. Use of new surface navigation technologies	1	2	3	4	5	N/A
6. Understand FAR Part 139 and Advisory Circular Safety Regulations	1	2	3	4	5	N/A
KNOWLEDGE INCREASE	None	Basic	Good	Sound	Expert	
7. My average level of knowledge and skill on movement area driving **before** completing driver training was:	1	2	3	4	5	
8. My average level of knowledge and skill on movement area driving **after** completing driver training was:	1	2	3	4	5	
ON-THE-JOB CONFIDENCE	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	N/A
After I completed the driver training, I was confident I would be able to:						
9. Understand the meaning of runway/taxiway signs, surface markings, and lighting	1	2	3	4	5	N/A
10. Understand ATC language, clearances, instructions, and light signals	1	2	3	4	5	N/A
11. Understand Notice to Airmen (NOTAMs)	1	2	3	4	5	N/A
12. Understand and use airport diagrams to navigate on the surface	1	2	3	4	5	N/A
13. Use of new surface navigation technologies	1	2	3	4	5	N/A

14. Relate FAR Part 139 and Advisory Circular Safety Regulations to driving on the airport surface	1	2	3	4	5	N/A
EFFECTIVENESS OF MATERIALS & METHODS	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	**N/A**
15. The driver-training program was clear and easy to follow.	1	2	3	4	5	N/A
16. The materials used provided enough information to facilitate the learning objectives	1	2	3	4	5	N/A
17. The overheads used are clear and easy to follow	1	2	3	4	5	N/A
18. The driver training video used is successful in illustrating the learning objectives	1	2	3	4	5	N/A
19. The exercises effectively cover the learning objectives	1	2	3	4	5	N/A
Type of Training						

20. This is traditional training (not computer interactive based training) Yes No (Circle)

21. This is AAAE interactive computer-based training? Yes No (Circle)

Demographic Information

22. Race (Circle One)
White – 1 African-American – 2 Hispanic – 3
Asian- Pacific Islander – 4 Native American – 5

23. Age (Circle One)
18yrs -25yrs – 1 25yrs – 35yrs – 2 35yrs – 45yrs – 3
45yrs – 55yrs – 4 55+yrs – 5

24. Education (Circle One)
No High School – 1 High School/GED – 2 Some College – 3
2yr College Degree – 4 4yr College Degree – 5 Master Degree – 6
Doctoral Degree – 7 Professional Degree JD, MD – 8

25. Income (Circle One)
20k or less - 1 20k – 30k – 2 30k – 40k – 3 40k – 50K
50k+ - 5

26. Martial (Circle One)
Single – 1 Married – 2 Separated – 3 Divorced – 4 Widowed – 5

Comments

1. What did you find **most** valuable about your driver training program? Please indicate why.
2. What did you find **least** valuable about your driver program? Please indicate why.
3. What improvements could be made to make your driver training program more effective?
4. The obstacles that stand in the way of the successful application of the knowledge and skills learned in this program are:
5. Overall Comments:

Appendix B

Multivariate Analysis of Variance Results for Significance of Association between Intermediate Independent Variables, and Training Methods across All Classes of Runway Incursions

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial ETA Squared
Intercept	Pillai's Trace	.043	4.275(a)	4.000	382.000	.002	.043
	Wilks' Lambda	.957	4.275(a)	4.000	382.000	.002	.043
	Hotelling's Trace	.045	4.275(a)	4.000	382.000	.002	.043
	Roy's Largest Root	.045	4.275(a)	4.000	382.000	.002	.043
Learning Objectives Met							
Understand the meaning of runway/taxiway signs, surface markings, and lighting	Pillai's Trace	.008	.778(a)	4.000	382.000	.540	.008
	Wilks' Lambda	.992	.778(a)	4.000	382.000	.540	.008
	Hotelling's Trace	.008	.778(a)	4.000	382.000	.540	.008
	Roy's Largest Root	.008	.778(a)	4.000	382.000	.540	.008
Understand ATC language, clearances, instructions, and light signals	Pillai's Trace	.025	2.442(a)	4.000	382.000	.046	.025
	Wilks' Lambda	.975	2.442(a)	4.000	382.000	.046	.025
	Hotelling's Trace	.026	2.442(a)	4.000	382.000	.046	.025
	Roy's Largest Root	.026	2.442(a)	4.000	382.000	.046	.025
Understand Notice to Airmen (NOTAMs)	Pillai's Trace	.018	1.759(a)	4.000	382.000	.136	.018
	Wilks' Lambda	.982	1.759(a)	4.000	382.000	.136	.018
	Hotelling's Trace	.018	1.759(a)	4.000	382.000	.136	.018
	Roy's Largest Root	.018	1.759(a)	4.000	382.000	.136	.018
Understand and use airport diagrams to navigate on the surface	Pillai's Trace	.015	1.500(a)	4.000	382.000	.202	.015

	Wilks' Lambda	.985	1.500(a)	4.000	382.000	.202	.015
	Hotelling's Trace	.016	1.500(a)	4.000	382.000	.202	.015
	Roy's Largest Root	.016	1.500(a)	4.000	382.000	.202	.015
Use of new surface navigation technologies	Pillai's Trace	.008	.734(a)	4.000	382.000	.569	.008
	Wilks' Lambda	.992	.734(a)	4.000	382.000	.569	.008
	Hotelling's Trace	.008	.734(a)	4.000	382.000	.569	.008
	Roy's Largest Root	.008	.734(a)	4.000	382.000	.569	.008
Understand FAR Part 139 and Advisory Circular Safety Regulations	Pillai's Trace	.016	1.580(a)	4.000	382.000	.179	.016
	Wilks' Lambda	.984	1.580(a)	4.000	382.000	.179	.016
	Hotelling's Trace	.017	1.580(a)	4.000	382.000	.179	.016
	Roy's Largest Root	.017	1.580(a)	4.000	382.000	.179	.016
Knowledge Increase							
My average level of knowledge and skill on movement area driving **before** completing driver training	Pillai's Trace	.047	4.664(a)	4.000	382.000	.001	.047
	Wilks' Lambda	.953	4.664(a)	4.000	382.000	.001	.047
	Hotelling's Trace	.049	4.664(a)	4.000	382.000	.001	.047
	Roy's Largest Root	.049	4.664(a)	4.000	382.000	.001	.047
My average level of knowledge and skill on movement area driving **after** completing driver training	Pillai's Trace	.032	3.109(a)	4.000	382.000	.015	.032

	Wilks' Lambda	.968	3.109(a)	4.000	382.000	.015	.032
	Hotelling's Trace	.033	3.109(a)	4.000	382.000	.015	.032
	Roy's Largest Root	.033	3.109(a)	4.000	382.000	.015	.032
On-the-Job Confidence							
Understand the meaning of runway/taxiway signs, surface markings, and lighting	Pillai's Trace	.024	2.355(a)	4.000	382.000	.053	.024
	Wilks' Lambda	.976	2.355(a)	4.000	382.000	.053	.024
	Hotelling's Trace	.025	2.355(a)	4.000	382.000	.053	.024
	Roy's Largest Root	.025	2.355(a)	4.000	382.000	.053	.024
Understand ATC language, clearances, instructions, and light signals	Pillai's Trace	.011	1.101(a)	4.000	382.000	.356	.011
	Wilks' Lambda	.989	1.101(a)	4.000	382.000	.356	.011
	Hotelling's Trace	.012	1.101(a)	4.000	382.000	.356	.011
	Roy's Largest Root	.012	1.101(a)	4.000	382.000	.356	.011
Understand Notice to Airmen (NOTAMs)	Pillai's Trace	.042	4.168(a)	4.000	382.000	.003	.042
	Wilks' Lambda	.958	4.168(a)	4.000	382.000	.003	.042
	Hotelling's Trace	.044	4.168(a)	4.000	382.000	.003	.042
	Roy's Largest Root	.044	4.168(a)	4.000	382.000	.003	.042

Understand and use airport diagrams to navigate on the surface	Pillai's Trace	.013	1.243(a)	4.000	382.000	.292	.013
	Wilks' Lambda	.987	1.243(a)	4.000	382.000	.292	.013
	Hotelling's Trace	.013	1.243(a)	4.000	382.000	.292	.013
	Roy's Largest Root	.013	1.243(a)	4.000	382.000	.292	.013
Use of new surface navigation technologies	Pillai's Trace	.019	1.860(a)	4.000	382.000	.117	.019
	Wilks' Lambda	.981	1.860(a)	4.000	382.000	.117	.019
	Hotelling's Trace	.019	1.860(a)	4.000	382.000	.117	.019
	Roy's Largest Root	.019	1.860(a)	4.000	382.000	.117	.019
Advisory Circular Safety Regulations to driving on the airport surface	Pillai's Trace	.004	.345(a)	4.000	382.000	.847	.004
	Wilks' Lambda	.996	.345(a)	4.000	382.000	.847	.004
	Hotelling's Trace	.004	.345(a)	4.000	382.000	.847	.004
	Roy's Largest Root	.004	.345(a)	4.000	382.000	.847	.004

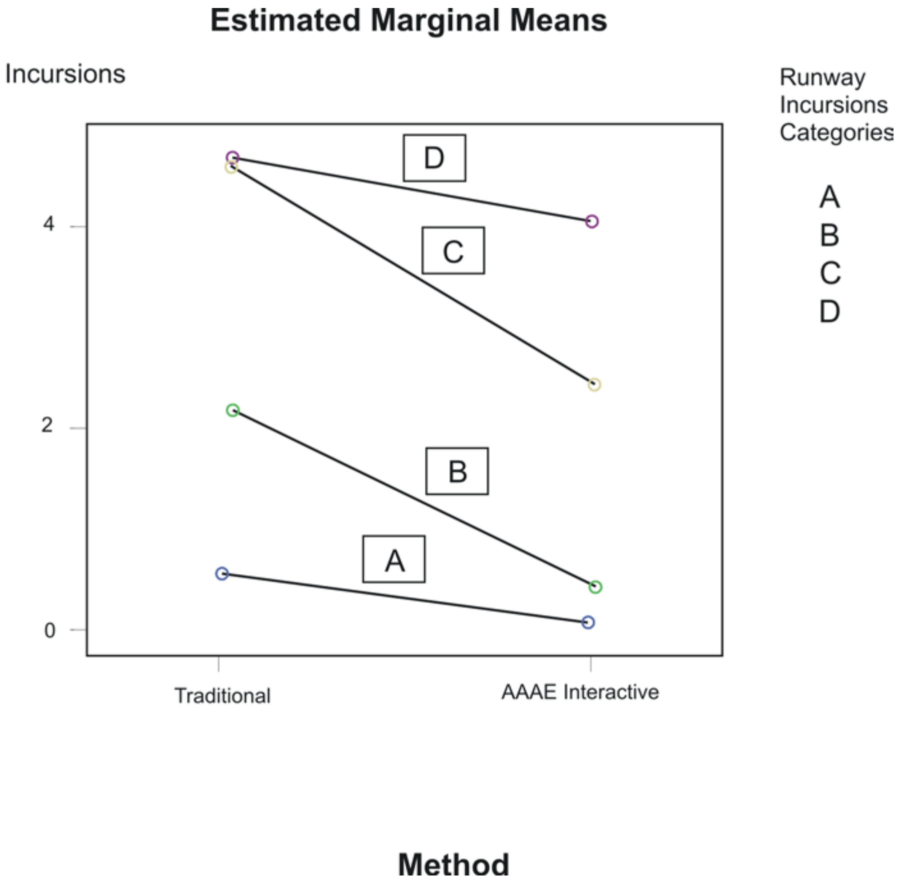
Effectiveness of Materials & Methods							
The driver-training program was clear and easy to follow	Pillai's Trace	.017	1.635(a)	4.000	382.000	.165	.017
	Wilks' Lambda	.983	1.635(a)	4.000	382.000	.165	.017
	Hotelling's Trace	.017	1.635(a)	4.000	382.000	.165	.017
	Roy's Largest Root	.017	1.635(a)	4.000	382.000	.165	.017
The materials used provided enough information to facilitate the learning objectives	Pillai's Trace	.079	8.213(a)	4.000	382.000	.000	.079
	Wilks' Lambda	.921	8.213(a)	4.000	382.000	.000	.079
	Hotelling's Trace	.086	8.213(a)	4.000	382.000	.000	.079
	Roy's Largest Root	.086	8.213(a)	4.000	382.000	.000	.079
The overheads used are clear and easy to follow	Pillai's Trace	.210	25.363(a)	4.000	382.000	.000	.210
	Wilks' Lambda	.790	25.363(a)	4.000	382.000	.000	.210
	Hotelling's Trace	.266	25.363(a)	4.000	382.000	.000	.210
	Roy's Largest Root	.266	25.363(a)	4.000	382.000	.000	.210

The driver training video used is successful in illustrating the learning objectives	Pillai’s Trace	.036	3.519(a)	4.000	382.000	.008	.036
	Wilks’ Lambda	.964	3.519(a)	4.000	382.000	.008	.036
	Hotelling’s Trace	.037	3.519(a)	4.000	382.000	.008	.036
	Roy’s Largest Root	.037	3.519(a)	4.000	382.000	.008	.036
The exercises effectively cover the learning objectives	Pillai’s Trace	.013	1.217(a)	4.000	382.000	.303	.013
	Wilks’ Lambda	.987	1.217(a)	4.000	382.000	.303	.013
	Hotelling’s Trace	.013	1.217(a)	4.000	382.000	.303	.013
	Roy’s Largest Root	.013	1.217(a)	4.000	382.000	.303	.013
Type of Training							
Method of Training	Pillai’s Trace	.191	22.515(a)	4.000	382.000	.000	.191
	Wilks’ Lambda	.809	22.515(a)	4.000	382.000	.000	.191
	Hotelling’s Trace	.236	22.515(a)	4.000	382.000	.000	.191
	Roy’s Largest Root	.236	22.515(a)	4.000	382.000	.000	.191

NOTE: a. Exact static.

Appendix C

Estimated Marginal Means – Profile Plots



International Journal of Applied Aviation Studies, Volume 7, Number 1

Developmental Reports

Analysis of Regional Oversight Organization as the Ultimate Form of Cooperation in Aviation Safety and Security Oversight

Vahid Motevalli
Aviation Institute
The George Washington University
20101 Academic Way
Ashburn, VA 20147
vahidm@gwu.edu

Abstract

Audit of civil aviation authorities have revealed areas of deficiencies. There have been an increasing number of initiatives to develop cooperative frameworks between States to address aviation safety oversight deficiencies. The spectrum of cooperation can range from resource sharing and partnership to bilateral and multi-lateral agreements. Regional Oversight Organization (ROO) is clearly the ultimate form of cooperation among civil aviation authorities requiring the greatest level of delegation of sovereignty in oversight. This paper reviews different mechanisms of cooperation and reasons for regional oversight organizations becoming more attractive to Civil Aviation Authority (CAAs). It also examines critical issues associated with a framework for developing, sustaining, and making effective an oversight organization with a regional scope.

Introduction

According to International Civil Aviation Organization (ICAO), an analysis of the results of the Universal Safety Oversight Audit Program (USOAP) indicated serious safety oversight deficiencies in a number of Contracting States and a lack of satisfactory progress in the resolution of safety concerns identified at the time of their initial audits (ICAO, 2004).

This experience has been further validated by the audit follow-up visits that indicated that while most Contracting States continued to make progress in the implementation of their corrective action plans and the resolution of safety issues, approximately 25 percent continue to experience difficulties in the implementation of their corrective action plans. This problem is not limited to any particular region or level of economic development (ICAO, 2004, p.1).

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

Return to Table of Contents

The last statement may not be intuitively obvious to most aviation professionals who track safety oversight issues.

The USOAP program has evolved in the last few years. ICAO no longer simply audits annexes 1, 6, and 8 but rather uses a comprehensive systems approach, which covers all annexes except annexes 9 and 17. Furthermore, full reports of the USOAP audits will be available to ICAO Member States as a change of policy. This policy change was approved in 2005, and was emphasized by Director Generals as a global strategy (ICAO, 2006).

Following such findings, ICAO proposed the Unified Strategy, which was adopted in the 35th Assembly with regional cooperation as a cornerstone. In addition, Document 9734b has been developed as a safety oversight manual for the establishment and management of regional safety oversight systems. However, ICAO being constrained as a consensus-making body is not able to forcefully recommend any single approach for regional safety oversight cooperation as being more effective. To some extent, there are limitations and regional, political, and economic considerations that make arriving at a single model rather difficult (FAA, 2005). The available approaches range from bilateral cooperation to formation of a regional authority. Clearly, the latter is only feasible for very small countries with cohesive regional interests or countries that have formed political union or federation. In general, the spectrum of cooperation among countries can be categorized into the following:

- Bilateral and Multi-lateral agreements
- Cooperative development of Operational Safety and Continuing Airworthiness Program (COSCAP)
- Regional Aviation Safety Oversight Organizations
- Regional Authority

Background

ICAO is a consensus making body. It must strike a balance between meeting the needs of developed States and those of the developing world. Therefore, the Annexes to the Convention establish only the minimal Standards and Recommended Practices (SARPs) for civil aviation authorities to conduct their oversight obligations. Although these are only minimal standards, complying with the standards can be difficult for States that do not have the necessary financial, technical, or human resources. Some States, which have the resources to devote to aviation, are finding it beneficial to cooperate with their neighbors and work together towards meeting their responsibilities under the Chicago Convention.

Historically, regional cooperation in aviation, among Member States, started due to necessity and their obligations to facilitate air navigation in air traffic control and provide search and rescue in case of aviation accidents. Also, meteorological services are other natural area for cooperation. Establishment of the Central American Corporation for Air Navigation Services (COCESNA) in 1960 to provide air traffic control facilitation in the region is a prime example. In this case, investments required by the Member States were radar facilities and Air Traffic Control (ATC). Other similar cooperative arrangements, mostly less structured, have taken place for ATC with Eurocontrol being one of the more sophisticated approaches. ICAO also initiated a number of regional cooperative programs under the COSCAP.

There are ten COSCAPs currently in existence with varying degrees of credible activities and funding, see Table 1 (Aviation Week and Space Technology, 2005).

Definitions

1. Regional Authority (RA) – Formation of a Regional Civil Aviation Authority by two or more countries via multi-lateral agreement and modification of their individual civil aviation law. Common regulations must be in place for all the States. The authority would represent all the States uniformly in aviation safety oversight and would act as a CAA for all members.
2. Regional Oversight Organization (ROO) – An organization formed by two or more countries with specific oversight responsibilities and enforcement abilities. The latter is desired but not essential. Each country retains their individual CAA, but there would be a need for varying degrees of harmonization of laws and regulations and modifications to the laws to allow this ROO to have the legal authority. Sharing of resources, responsibilities, and enforcement authority is agreed to by the member States.
3. Regional Cooperation – No organization, but cooperative mechanisms developed via multilateral agreements. State laws and regulations are not modified.

Table 1
COSCAP Regional Organizations and their Characteristics

Project	Participating States	Duration	Cost
Former CIS States	Armenia, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Azerbaijan, Tajikistan	2002-06	$3 M
West Africa (UEMOA)	Benin, Burkina Faso, Cote d’Ivoire, Guinee Bissau, Mali, Niger, Senegal, Togo, Mauritania	2003-05	*
South Africa (CEMAC)	Central African Republic, Cameroon, Congo, Equatorial Guinea, Gabon, Chad, Sao Tome & Principe	2003-05	*
West Africa (Banjul Accord)	Cape Verde, Gambia, Ghana, Guinea, Liberia, Nigeria, Sierra Leone	2003-05	*
East Africa (EAC)	Kenya, Tanzania, Uganda	*	*
Latin America	Argentina, Bolivia, Brasil, Chile, Cuba, Ecuador, Panamá, Paraguay, Uruguay, Perú	2001-06	$2.6M

South Asia	Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka	1998-07	$5 M
South East Asia	Brunei, Cambodia, Hong Kong, Macau, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Viet Nam, South Korea	2002-06	$1.5M
North Asia	China, Japan, North Korea, South Korea, Mongolia	2003-07	$2 M
Pacific Aviation Safety Office (PASO)	Australia, Cook Islands, Fiji, Kiribati, Marshall Islands, Micronesia, Nauru, New Zealand, Niue, Palau, Papua, New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu	*	*

Note: From “ICAO COSCAP Initiatives,” by ICAO, 2004. * = no data. CIS = Commonwealth of Independent States, UEMOA = West African Economic and Monetary Union, CEMAC = Economic and Monetary Community of Central Africa, EAC = East African Community. EAC is moving to form a regional oversight organization.

Competing Priorities: Sovereignty vs. Efficiency

At the core of decision to engage in a regional cooperative initiative is the willingness, along with political, economic, and technical feasibility of the States, to balance the efficiencies of scale and scope in cooperation. The extent of cooperation must be weighed versus the degree of willingness to relinquish oversight rights, which translates into sovereignty. Figure 1 is a graphical depiction of this balance. A Regional Authority (RA), such as that established by the OECS (Organization of Eastern Caribbean States) has a complete shared authority on behalf of the individual States, while COSCAPs are much closer to an individual State CAA in terms of sharing their oversight authority. The European Aviation Safety Agency (EASA), empowered by the European Union (EU) was established in 2002, and is certainly a ROO, and in parts approaches a Regional Authority. On the other hand, the Central American Aviation Safety Agency (ACSA), established in 1999, is regarded as a success story of cooperation (Jennison, 2006; GW Aviation Institute, 2005) is a good model of Regional Oversight Organizations (ROO). ACSA and others, such as Regional Aviation Safety Oversight System (RASOS), represent degrees of shared authority and responsibility fitting the model of a ROO to a different extent and are defined by the multi-lateral agreements. A more recent development is the East African Community (EAC) aviation safety authority that is being developed by Kenya, Tanzania, and Uganda and is expected to be in place by January of 2007. The laws in all three EAC countries allow for such a development and their regulations are being harmonized based on Kenyan regulations, which were chosen as the baseline. Additionally, radar training facilities and standard operating procedures for rotation of personnel have been developed.

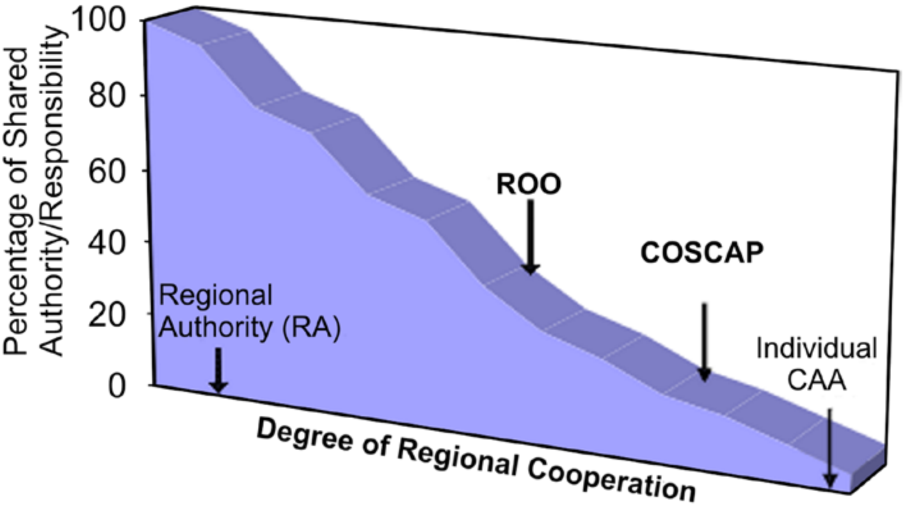


Figure 1. Depiction of variation in regional cooperation.

Establishing a Regional Oversight Organization

Establishment of a ROO does not exempt a Contracting State from its obligations under the Chicago Convention and its Annexes. However, it allows for a more thorough and complete compliance and, if implemented well, an improved safety oversight system. Overarching issues concerning the development of a regional cooperation initiative including the following:

- Impact on State's sovereignty
- Political commitment at the national level to enable compromises on sovereignty issues and to provide the requisite resources
- A thorough understanding of the economic benefits gained from safety oversight improvements and development of an aviation system, and
- A clear understanding of what obligations must be retained directly by the State

The economic benefits for ROO members are often expressed in terms of leveraging economies of scale and scope. Perhaps the more useful terminology is *efficiency* of scale and scope. For compelling reasons, economic benefits of ensuring safety and security oversight, ROOs can provide improvements due to both efficiencies and economics of scale and scope.

Regional organizations' ultimate objective should be enabling full compliance with ICAO standards by *all* of their Member States. Full compliance does not mean meeting the minimum Standards established by the Annexes, but rather compliance with the spirit intended by these Annexes. In practice, partnerships should encourage States to assist neighbors (not necessarily a geographic relationship) with preparations for regular safety and security audits. Other regional oversight organizations have been formed outside the COSCAP organizations, including the Central American Aviation Safety Agency (ACSA). ACSA was formed in 2000 out of the Central American Air Navigation Services, an international governmental corporation providing air navigation services (GW Aviation Institute, 2005). ACSA was charged with providing safety oversight services to its Member States, including Belize, Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua. Its priorities include:

- Compliance with the ICAO standards;
- Ensuring uniform levels of safety among member States (based on the Eight Critical Elements of safety oversight – ICAO, 2005), by establishing standardized safety oversight systems and processes for the region;
- Eliminating duplication;
- Establishing a cost effective system of safety oversight.

The member States benefit from economies of scale such as sharing the cost of highly qualified technical personnel and developing commercial activity in the region.

Efficiencies of Scale and Scope

It is intuitively clear that better efficiencies may be obtained in a larger scale system as resources, on average, can be utilized more effectively. In particular, sustaining the level of effort for qualified technical personnel is one of the most challenging aspects for an oversight organization. Sustaining the level of effort can be shown to benefit from efficiencies of scale. For example, a single certified operations inspector for a particular aircraft type may be underutilized in a country where there are only two such aircrafts. The same inspector would be adequately utilized within a regional organization, which may contain four to five aircraft of that type among the Member States. Another example of efficiency of scope may involve software resources, such as a safety oversight database or training programs. By pooling resources, better products and programs with much more extensive scope can by developed for per capita lower cost.

Success of Regional Oversight Organizations

Success of regional organizations in many fields, particularly those anchored in economic treaties and unions, is well known. A regional organization for the sole purpose of oversight brings an entirely different aspect to the collaboration. For example, a regional organization established for trade development has many similar components as an ROO, such as harmonization of laws and regulations and requiring qualified personnel and shared responsibilities. However, key differences would be the extent of compromise on sovereignty issues and metrics of success. For a ROO, the first is required and the outcomes are often more difficult to measure, but a regional organization for trade may not necessarily need to relinquish on sovereignty issues and economic outcomes are a lot easier to measure. Development of Regional Oversight Organizations is a concrete and profound step for any government to address their obligations under ICAO Conventions and the Annexes to the Conventions. A Regional Oversight Organization is a well-defined approach among different mechanisms for cooperation among Member States. The success of this approach may depend on three specific issues:

- START WITH VERB Specified details for the development of the Regional Oversight Organization
- Ensuring revenue streams, infrastructural, and political support to sustain the organization
- Devising the necessary mechanisms and tools to make the organization effective

The fundamental issues affecting the success of a ROO are described as follows:

For the *development* of the Regional Oversight Organization, the following points need to be considered and acted on:
(a) Political commitment at the highest levels of each Member State is required
(b) A foundation in the law of the country must exist or be developed to address both internal and multilateral issues and be ratified by respective legislative bodies (See ICAO 9734A, 2005 and 9734B, 2006)
(c) An initial investment of resources and funding is required
(d) Such an organization can greatly benefit from existing economic/political structure and treaties and suitable venues should be explored
(e) Need a leader, local champion to pursue the development of the ROO.

Making the Regional Oversight Organization *effective* requires the following key issues to be covered:
(a) Adequate and qualified technical personnel;
(b) Dynamic regulation and enforcement;
(c) Institutionalization of organization's continuity, and
(d) A strategic framework and a management system to ensure long-term success.

Sustaining a Regional Oversight Organization requires the following essential elements:
(a) Annual reliable budget;
(b) Development of dedicated sustainable revenue sources, and
(c) Recruitment and retention of technical and administrative resources.

These issues are not unlike those facing individual Civil Aviation Authorities and are equally challenging to establish. The advantage is that if a number of States agree to form a regional organization, they can address these issues for their individual States as well as the regional organization at the onset and plan for their implementation.

Critical Elements of Safety Oversight

The use of eight critical elements of safety oversight as outlined in the ICAO 9734 Document in audits programs, particularly under IASA and USOAP programs, is well established. A key question in forming an ROO would be how well can the eight critical areas of safety oversight be mapped onto a ROO? Figure 2, attempts to show how each of these areas can be mapped and highlights the advantages for a ROO.

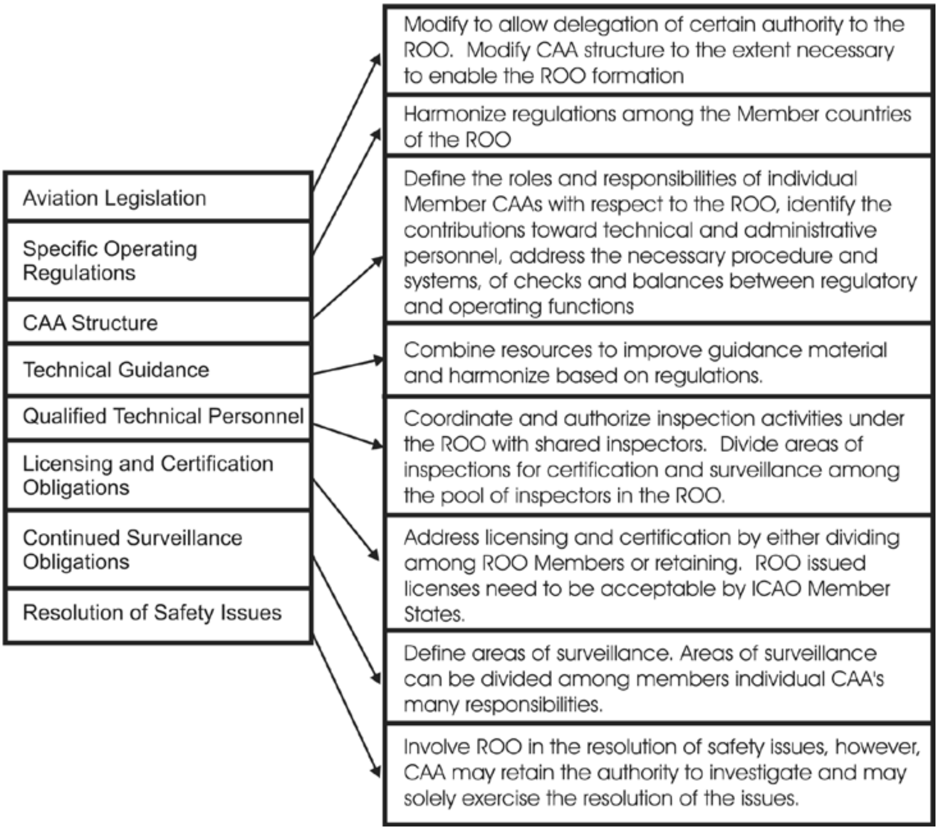


Figure 2. Mapping the eight critical elements of safety oversight onto a Regional Oversight Organization.

A model ROO can be built on strong political commitment from the national governments to ensure an autonomous status and adequate resources. Adequate changes in the laws of the member countries can establish and ensure that a ROO performs oversight functions while minimizing complicated and intractable political debate. The autonomy of the ROO will allow it to complete its regulatory mission with a director and organization that is free from political pressure. Ideal characteristics of this independence include; the legal authority to penalize, a separate budget, a board of directors comprised of ministerial level individuals from each country, a direct relationship between the budget of the ROO and its performance, and aviation activities that generate revenue.

Adequate technical resources form an integral part of a ROO because the organization must have the necessary technological advantages in order to ensure effective and efficient oversight of the complex aviation industry. Given the advancements of technology and their effect on regulations and due to natural turnover rates of qualified personnel, a ROO should have the capability to provide adequate training. Human resources include a large enough pool of qualified technical personnel in the fields such as general management, database management, legislation, security, airworthiness, operations, accident investigations, air traffic management, and airport operations.

For those countries that have the political commitment, but lack the adequate resources, the author suggests that regional cooperation will help each of their CAAs to attain the necessary resources and that a ROO may ultimately be needed to achieve the required oversight functions.

Evidence of Success

Audit of civil aviation organizations by ICAO (under the USOAP program) or by other Member States, such as the IASA program in the U.S. or the EASA audits by the EU, provide direct feedback on the compliance of a Member State with ICAO Annexes. In addition, IATA audits of airlines provide additional opportunity to assess the safety of the system and indirectly, the oversight and regulatory effectiveness. However, this information is not shared publicly, although the ICAO 36th Assembly has adopted a resolution to make these audit reports available. In lieu of detailed data, general information can be used to assess the effectiveness of oversight performance of countries. Table 2 shows de-identified compliance data from six countries for the *eight critical areas of safety oversight*. These countries, de-identified as the data were derived from confidential reports, were selected for the similarity of the size of their aviation industry and relative similarities of economic strength and stability. The table shows the percent of change in effective implementation of ICAO Annexes 1, 6, and 8. The countries numbered in bold are members of ROOs. In general, this data shows that a follow-up audit revealed significant improvement in compliance by states in an ROO and in most of the critical areas of safety oversight. The areas indicated in blue show compliance of 95% or better. In contrast, those countries not being involved with an ROO (1, 2, and 6) show either very little improvement or not adequate improvement over the course of a follow-up audit. In some cases, where indicated by orange color, the lack of effective implementation (LEI) is below the global average at the follow-up audit.

Table 2

Percent changes in compliance with ICAO Annexes and SARPs. Percentage is derived from follow-up audits compared to the original audit and is organized by the eight critical areas of safety oversight

Country	Primary Aviation Legislation	Operat-ing Regs	CAA Structure & Safety Oversight Functions	Tech Guidance Material	Qualified Tech Personnel	Licensing & Certification	Continued Surveil-lance	Resolution Of Safety Issues
1	3	🟧 0	🟧 2	🟧 0	4	2	🟧 0	0
2	23	🟧 -7	7	🟦 37	25	25	35	80
3	🟦 30	10	🟧 7	15	🟧 17	16	4	30
4	50	🟦 50	🟦 30	🟦 65	🟦 52	43	37	🟦 70
5	🟦 5	🟦 7	🟦 5	🟦 15	18	🟦 14	🟦 7	🟦 50
6	🟦 5	🟦 15	7	35	15	8	15	🟦 20

🟧 Below global LEI for specific category
🟦 LEI <= 5%

Utility, success, and effectiveness of the regional organization can be further demonstrated by examining effectiveness of States versus the ROO in each of the eight critical areas as shown in Figure 3. The data shows average LEI and EI (effective implementation) for individual States vs. ROO member States. As expected, the areas of qualified technical personnel and licensing and certification are two areas with highest LEI. However, the same data shows that use of an ROO significantly contributes to the effective compliance with the qualified technical personnel area. Figure 3 also shows that a ROO can contribute significantly to improved compliance in seven of the critical areas while compliance with requirements in law and legislation is best accomplished by individual States and a ROO cannot contribute significantly to this area.

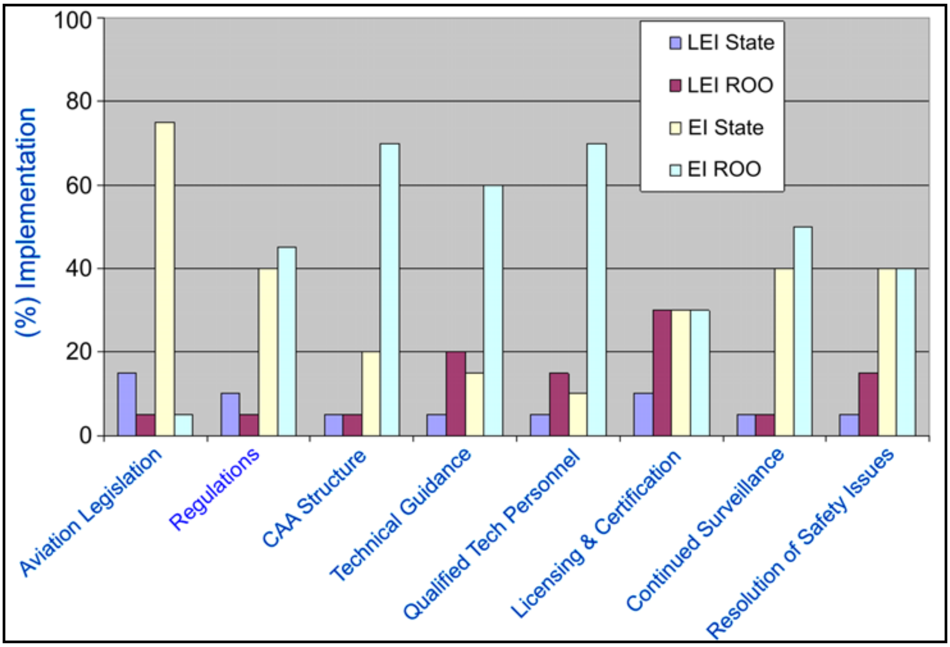


Figure 3. Comparison of effective implementation of the eight critical elements for individual States vs. ROOs.

Conclusions

The discussion in this paper describes the advantages in developing a regional oversight organization as opposed to limited cooperation among States in a region through bilateral and multi-lateral agreements. Detailed discussions of the essential elements for developing, sustaining, and the effectiveness of the regional oversight organizations have been outlined here. Based on the presentation of this limited data and discussion of the essential elements, it is clear that a regional oversight organization can benefit from efficiencies of scale and scope. Although compromises on sovereignty may seem unthinkable, they have been achieved by many countries to advance safety oversight and the outcome has been acceptable. The following recommendations and concluding thoughts are offered:

(a) International programs for outreach, training, technology transfer and harmonization should be focused around regional cooperation

(b) ICAO needs to develop a focused and coordinated effort to provide efficient and streamlined support for developing nations and regional organizations are one of the best vehicles for many States to comply with their ICAO obligations, and

(c) A working model/laboratory for oversight tools and regional organizations should be developed by FAA or ICAO.

References

Aviation Week & Space Technology, (2005, August). *163*(8), 77.

Federal Aviation Administration. (2005, October 20). Breakout Session Panel A: Global Aviation in the 21st Century is Nearly a Seamless Operations. *2nd Annual FAA International Aviation Safety Forum.*

GW Aviation Institute. (2005). *Reports on International Summits on Aviation Safety and Security.*

International Civil Aviation Organization. (2006, March) Directors General of Civil Aviation, *Conference on a Global Strategy for Aviation Safety.*

International Civil Aviation Organization. (2006). Safety *Oversight Manual - Part B: The Establishment and Management of a Regional Safety Oversight System* (Doc 9734-AN/959, First Edition)

International Civil Aviation Organization. (2005) *Safety oversight manual - Part A: The establishment and management of a state's safety oversight system* (Doc 9734-AN/959, Second Edition).

International Civil Aviation Organization. (2004). Working Paper, Item 16. *35th Assembly*.

Jennison, M. (2006). Regional safety oversight bodies deliver economies of scale and greater uniformity, ICAO Journal. *61*(1), 9-12.