Office of the Chief Scientist for Human Factors

General Aviation Human Factors

Program Review
FY05

William K. Krebs Ph.D., Aviation Maintenance Human Factors Program Manager

Federal Aviation Administration
ATO-P R&D HF (Room 907A)
800 Independence Avenue, S.W.
Washington, D.C. 20591
phone (202) 267-8758
e-mail: william.krebs@faa.gov
http://www.hf.faa.gov/krebs

The Federal Aviation Administration Office of the Chief Scientific and Technical Advisor for Human Factors (ATO-P R&D HF) directs a general aviation research program that focuses on reducing fatalities, accidents, and incidents within the general aviation flight environment. This environment is defined as all flights that are conducted under FAR Part 91 as well as the general aviation maintenance community. The research addresses better methods for the detection, classification, and reporting of human factors accidents; developing certification and flight standards and guidelines based on human factors research, and identifying and implementing intervention strategies to impact general aviation accidents.

The following report summarizes projects between October 1st, 2004 and September 30th, 2005. These projects attempt to address requirements identified by the Federal Aviation Administration Flight Standards and Certification offices. The intent of this report is to allow Federal Aviation Administration sponsors to determine whether their requirements have been satisfactorily addressed, allow investigators to receive feedback from Federal Aviation Administration sponsors and other interested parties, and to provide feedback to the ATO-P R&D HF general aviation program manager on the quality of the research program. Basically, this document is a means of holding each group (sponsor, investigator, ATO-P R&D HF program manager) accountable to ensure that the program is successful.

In FY05, the general aviation research program distributed $437,000 contract dollars to performing organizations. In addition, some of these projects received supplemental support from the Civil Aerospace Medical Institute, Oklahoma City, OK.

Address questions or comments to:

William K. Krebs, Ph.D.

Project Title	Requirement Statement	Sponsor	Research Requirement link
Aviation Safety Inspector (ASI) Training for Technically Advanced Aircraft	Recently, there has been an emergence of technically advanced (i.e., glass cockpit) aircraft (TAA) within general aviation. Aside from technical challenges presented by the design of these advanced avionics systems, there are difficulties in acquiring a conceptual understanding of the functions offered by the avionics, developing system monitoring skills and habits, developing mode management and awareness skills, understanding when and when not to use automation, and maintaining manual flying skills. Operating aircraft with advanced avionics requires an additional set of knowledge elements and skills. Currently, FAA aviation safety inspectors (ASIs) are required to inspect technically advanced aircraft, check certified flight instructors, and conduct surveillance of designated pilot examiners who are certifying pilots operating technically advanced aircraft. However, many of the aviation safety inspectors within the FAA workforce completed flight training prior to the entry of advanced avionics.	AFS-800	link
Developing a Methodology for Assessing Safety Programs Targeting Human Error in Aviation	Four the past five years, the Civil Aerospace Medical Institute and the University of Illinois have systematically examined over 20,000 general aviation accidents occurring between 1990-2000 using the Human Factors Analysis and Classification System (HFACS). HFACS is a theoretically based framework used widely throughout aviation and other high-risk industries for investigating and analyzing human error associated with accidents and incidents. The HFACS framework has been reliably used to analyze the underlying human factors causes of both commercial and general aviation accidents and has helped identify general trends in the types of human factors issues and aircrew errors that have contributed to civil aviation accidents. Previous HFACS research performed at both the University of Illinois and the Civil Aerospace Medical Institute (CAMI) has shown that HFACS can be reliably used to analyze the underlying human factors causes of both commercial and GA accidents (Wiegmann & Shappell, 2001, 2003; Shappell & Wiegmann, 2003). Furthermore, these analyses have helped identify general trends in the types of human error that have contributed to civil aviation accidents. For example, when the GA accidents between 1990-2000 were examined using HFACS, several heretofore unknown facts regarding GA aviation safety were revealed (Figure 1). It appears that safety efforts over the last	AFS-800	link

	several years have had little impact (flat trend lines) on any specific type of human error associated with GA accidents. If anything, they have had a ubiquitous impact – albeit unlikely. Equally noteworthy, skill-based errors have contributed to GA accidents more than any other error form (roughly 80% of all GA accidents examined). While data such as these are important, the next step in the process is to identify a variety of intervention strategies to either prevent or mitigate general aviation accidents. The purpose of this research is to do just that. However, rather than recycle or continue to employ the same old intervention strategies this requirement will address a new approach to the development of accident/incident prevention/mitigation.		
USE OF WEATHER INFORMATION BY GENERAL AVIATION PILOTS: PROVIDERS AND PRODUCTS	Weather is the single largest cause of aviation fatalities, especially in general aviation (GA). As part of a multi-pronged effort to understand why pilots continue to experience weather-related accidents and incidents, this requirement seeks to develop baseline information on how "typical" GA pilots acquire, evaluate, and use weather information. This proposed requirement addresses the agency's goal of reducing GA fatalities. It also supports interventions recommended by the GA Joint Steering Committee (GA-JSC), a government-industry group that oversees and tracks accident reduction efforts.	AFS-800	link
General Aviation Private Pilot Survey / Designated Pilot Examiner Program Assessment	The FAA's 2004-2008 Strategic Plan (Flight Plan) Increased Safety goal's objective two intends "to reduce the number of fatal accidents in general aviation." In order to meet this objective, the FAA's General Aviation and Commercial Division (AFS-800) plans to improve the Designated Pilot Examiner (DPE) program. The DPE administers a practical test to evaluate the examinee's (pilot) knowledge and skill to perform a task. When the DPE evaluates knowledge there should be 1) an adequate coverage of the knowledge domains, 2) consistency in the level of difficulty of questions across domains, 3) consistency in how the questions are presented, and 4) consistency in examiner knowledge of the goals of the examination. Problems arise when the DPEs are not consistent in the way they conduct practical tests. The variance could occur between examinees or within an examination – or between DPEs. Research objectives include: what kind of intervention strategies can be	AFS-800	link

	developed to improve DPE performance? What kind of intervention strategies can improve the quality of FAA oversight of examiners? How effective are the interventions? How prepared and organized are the DPE's when conducting practical tests? To accomplish the research objectives, a general aviation private pilot national survey (similar to FAA's ASW DPE general aviation private pilot survey) will be administered to all FSDOs within the United States. Newly certified general aviation private pilots should complete the survey within six weeks of the practical test. The FSDOs with the assistance of a contractor will provide a list of the GA pilots certified along with their home addresses to CAMI. National private pilots' will be surveyed in mid-2005. The purpose is to measure the effectiveness of the practical tests that examiners delivered in 2005.		
Symbol Set Discriminability Metrics: Extending Discrimination Models for Size and Position Invariance	Do the traffic symbols proposed in the draft Advisory Circular, "Aircraft Surveillance Systems and Applications" meet the basic human factors requirement of discriminability?The experiment must validate the discriminability of a set of surveillance traffic symbols. The experiment should focus on "low-end" displays: such as those that are have small size, low pixel pitch, etc. The symbols to be validated will be provided by FAA AIR-130, and will number approximately 20. This number will include the two basic shapes to indicate directionality (chevron and diamond), their proximity and alert status, and their selection status (selection by the flight crew to display additional information). Other information coding such as air/ground status and information quality status may also be explored, depending on the experimental resources available. The symbols in the experiment should also include those from TCAS, as well as other symbols (e.g., navigation) that have potential to be confused with traffic symbols. The experiment should only address discriminability. One method might be to display one symbol at a time to the human subject, who can then demonstrate discriminability by identifying that symbol on a fixed, master symbol list that contains all symbols used in the experiment. Because of the limited scope of this experiment, it is not necessary to present symbols in a flight deck context. Furthermore, it is not necessary to use pilots for human subjects. However, it is important that all other experimental conditions are chosen such that the results from this limited experiment provide a meaningful validation of	AIR-130	link

	traffic symbol discriminability that can be applied to actual flight conditions.Control variables could potentially include symbol size, symbol luminance, symbol color contrast, pixels-per-symbol, symbol rendering method, display pixel pitch, display brightness, ambient lighting, and display pixel size. The majority of these and other parameters will be held constant throughout the experiment. However, one or more of these parameters will need to be varied to construct two trials that are designed to yield statistically significant differences in symbol discrimination performance. The experiment will consist of two trials: 1) a representative "realistic" scenario for low-end displays that is expected to yield a low error rate, and 2) an improbable "difficult" scenario that is expected to yield a higher error rate. The purpose of the difficult scenario is to establish some sense of what would *not* be acceptable to the FAA. It would also provide valuable data for future experiments that investigate the mechanisms of symbol discriminability, including the validation of software discrimination tools. Prior to the main experiment, a preliminary experiment will be to determine or adjust conditions such as precise symbol attributes (size, rendering, color, shape), lighting, performance estimates, methodology, etc. The goal during the preliminary experiment is to determine with sufficient certainty, and in a limited amount of time, the detailed conditions that are most appropriate for the experiment. Some degree of prototyping and iteration may be required. After the main experiment is complete, the data must clearly establish validation of the symbol set. The validation criteria should be based not only on the number of discrimination errors, but also on the mechanism of these errors, which will likely be rare events. Such events may be investigated in real time, if appropriate, in order to better determine the mechanism underlying discrimination errors. The criteria for validation should largely be determined prior to any prototyping and experimentation, with slight adjustments acceptable during the preliminary phase to account for discrimination errors that are an artifact of the experimental method.		
ENHANCED GENERAL AVIATION DECISION	A number of people from industry, academia, and the Federal Aviation Administration believe that the general aviation training programs do not have the flexibility to adapt to the wide variety of aviation technology (e.g., GPS, multifunction displays with moving map navigation, and traffic, weather, and	AFS-800	link

MAKING: WEATHER ASSESSMENT DURING PREFLIGHT PLANNING	terrain avoidance systems) that has recently emerged in the national airspace. With older technology systems, it did not matter who built the system since they all functioned and looked similar. However, with new technology, systems that perform similar functions may not look alike and pilot interaction with these systems may be completely different. Consequently, a "one-size-fits-all" approach to training may no longer be adequate. FAA/Industry Training Standards (FITS) will attempt to overcome the limitations of existing training programs by working in collaboration with industry to develop new and innovative training methods to ensure that pilots are trained and maintain proficiency in aircraft that contain new technology. New training methods emphasize improved risk management, training and education, and proper use of new technology.1155		
Unmanned Aircraft Pilot Medical and Certification Requirements	To support the rapidly growing industry of Unmanned Aerial Vehicles (UAVs) as they transition from military to non-military surveillance and cargo applications. Historically, flight operations assumed an onboard pilot controlling an aircraft to ensure safe operation in the National Airspace System (NAS). An unmanned aerial vehicle may be controlled by a pilot or operator from a distant ground station, and in some cases, operate autonomously where the UAV's flight path is based on pre-programmed global position system waypoints and the ground pilot or operator has very limited control over the aircraft flight movements. UAVs offer exciting opportunities for civil aviation; however before non-military UAV operations are fully integrated into the NAS, the FAA's General Aviation and Commercial Aviation Division (AFS-800) needs to define operator qualification and training requirements.	AFS-800	link
HOW HIGH IS HIGH ENOUGH? QUANTIFYING THE IMPACT OF AIR TRAFFIC CONTROL TOWER OBSERVATION	The purpose of this project is to develop research and educational materials that will help reduce accidents caused by 4 related problems: 1) continued flight into reduced visibility, 2) failure to detect targets, 3) failure to utilize resources, 4) need for improved education and training for problems 1-3. A review of the current literature indicates that accidents related to visibility account for a large portion of the total fatalities in aircraft. Visibility issues range from continued flight into instrument meteorological conditions (IMC) resulting in controlled	AFS-800	link

HEIGHT ON DISTANCE PERCEPTION	flight into terrain (CFIT), runway incursions and ground-based accidents during low visibility conditions, and midair collisions with ground-based objects or other aircraft. These mid-air collisions are often due not only to reduced visibility, but also to background conditions that camouflage or mask the target and impede detection, and indeed many of these accidents occur in clear skies. In most situations there appears to be a failure on the part of the pilots to recognize unsafe visual conditions and take appropriate action. In addition, reports indicate that in many cases, pilots of accident aircraft did not avail themselves of available technology, either advanced equipment installed in the aircraft, or ATC services. Further research aimed at understanding visual limitations under conditions of low visibility and decreased detection is needed. Such research would include optimizing strategies for employing available technology and services. Results from this research will form the basis for education materials designed to improve pilot recognition and performance under non-optimal visual conditions, and ultimately reduce accidents related to poor visual conditions.1770		
Visibility in the Aviation Environment	The purpose of this project is to develop research and educational materials that will help reduce accidents caused by 4 related problems: 1) continued flight into reduced visibility, 2) failure to detect targets, 3) failure to utilize resources, 4) need for improved education and training for problems 1-3. A review of the current literature indicates that accidents related to visibility account for a large portion of the total fatalities in aircraft. Visibility issues range from continued flight into instrument meteorological conditions (IMC) resulting in controlled flight into terrain (CFIT), runway incursions and ground-based accidents during low visibility conditions, and midair collisions with ground-based objects or other aircraft. These mid-air collisions are often due not only to reduced visibility, but also to background conditions that camouflage or mask the target and impede detection, and indeed many of these accidents occur in clear skies. In most situations there appears to be a failure on the part of the pilots to recognize unsafe visual conditions and take appropriate action. In addition, reports indicate that in many cases, pilots of accident aircraft did not avail themselves of available technology, either advanced equipment installed in the aircraft, or ATC services. Further research aimed at understanding visual limitations under conditions of low visibility and decreased detection is needed. Such research would include	AFS-800	link

	optimizing strategies for employing available technology and services. Results from this research will form the basis for education materials designed to improve pilot recognition and performance under non-optimal visual conditions, and ultimately reduce accidents related to poor visual conditions.1770		
A Human Factors Analysis of General Aviation Accidents in Alaska Versus the Rest of the United States	The Human Factors Analysis and Classification System (HFACS) is a theoretically based tool for investigating and analyzing human error associated with aviation accidents and incidents. Previous HFACS research performed at both at the University of Illinois and the Civil Aerospace Medical Institute (CAMI) has been highly successful and has shown that HFACS can be reliably used to analyze the underlying human factors causes of both commercial and general aviation accidents. Furthermore, these analyses have helped identify general trends in the types of human factors issues and aircrew errors that have contributed to civil aviation accidents. Key members of the FAA (e.g., AFS-800) and several committees chartered to address general aviation safety (e.g., Aeronautical Decision Making (ADM) JSAT and the General Aviation Data Improvement Team (GADIT)) have acknowledged the added value and insights gleaned from these HFACS analyses. However, these individuals and committees have directly requested that additional analyses be done to answer specific questions about the exact nature of the human errors identified, particularly within the context of general aviation. The purpose of the proposed research project, therefore, is to address these questions by performing a more fine-grained HFACS analysis of the individual human causal factors associated with fatal GA accidents and to assist in the generation of possible intervention programs.1453	AFS-800	link
Human Error Associated with Air Medical Transport Accident in the United States	The Human Factors Analysis and Classification System (HFACS) is a theoretically based tool for investigating and analyzing human error associated with aviation accidents and incidents. Previous HFACS research performed at both at the University of Illinois and the Civil Aerospace Medical Institute (CAMI) has been highly successful and has shown that HFACS can be reliably used to analyze the underlying human factors causes of both commercial and general aviation accidents. Furthermore, these analyses have helped identify general trends in the	AFS-800	link

	types of human factors issues and aircrew errors that have contributed to civil aviation accidents. Key members of the FAA (e.g., AFS-800) and several committees chartered to address general aviation safety (e.g., Aeronautical Decision Making (ADM) JSAT and the General Aviation Data Improvement Team (GADIT)) have acknowledged the added value and insights gleaned from these HFACS analyses. However, these individuals and committees have directly requested that additional analyses be done to answer specific questions about the exact nature of the human errors identified, particularly within the context of general aviation. The purpose of the proposed research project, therefore, is to address these questions by performing a more fine-grained HFACS analysis of the individual human causal factors associated with fatal GA accidents and to assist in the generation of possible intervention programs.1453		
Human Error and Commercial Aviation Accidents: A Comprehensive, Fine-Grained Analysis Using HFACS	The Human Factors Analysis and Classification System (HFACS) is a theoretically based tool for investigating and analyzing human error associated with aviation accidents and incidents. Previous HFACS research performed at both at the University of Illinois and the Civil Aerospace Medical Institute (CAMI) has been highly successful and has shown that HFACS can be reliably used to analyze the underlying human factors causes of both commercial and general aviation accidents. Furthermore, these analyses have helped identify general trends in the types of human factors issues and aircrew errors that have contributed to civil aviation accidents. Key members of the FAA (e.g., AFS-800) and several committees chartered to address general aviation safety (e.g., Aeronautical Decision Making (ADM) JSAT and the General Aviation Data Improvement Team (GADIT)) have acknowledged the added value and insights gleaned from these HFACS analyses. However, these individuals and committees have directly requested that additional analyses be done to answer specific questions about the exact nature of the human errors identified, particularly within the context of general aviation.	AFS-800	link

	The purpose of the proposed research project, therefore, is to address these questions by performing a more fine-grained HFACS analysis of the individual human causal factors associated with fatal GA accidents and to assist in the generation of possible intervention programs.1453		
TRANSFER OF TRAINING EFFECTIVENESS OF A FLIGHT TRAINING DEVICE (FTD)	Modern flight training devices provide a more effective and safe training experience than aircraft. Instructor and student discuss, perform, and review specific maneuvers in a quiet environment, without the distractions of danger of other aircraft, weather, etc. FTDs provide emergency procedures often not possible in an aircraft. Further, the quality of flight training will be more uniform if the most credit is reserved for the most capable devices, and less credit granted for less capable machines. By adjusting the flight credit allowance per the varying capabilities of FTDs, the FAA shows that it recognizes qualitative differences in the training experience. It is anticipated that a regulation change may provide incentive for further FTD development and use, and an increase in training effectiveness and efficiency. SubTasks: 1. Evaluate all seven levels of FTDs, recategorizing them as necessary by shared characteristics (i.e., fidelity fo physical/visual/flight replication) 2. Develop a system for measuring and recording a range of pilot performance within the areas of aircraft handling, navigation, and emergency procedures. 3. Measure the performance levels of students from each of the seven FTD categories. 4. Determine the point at which performance levels in an aircraft meet pilot certification standards?	AFS-800	link

General Aviation Human Factors

FY05 Funded Projects

A human factors analysis of general aviation accidents in Alaska versus the rest of the United States

Cristy Detwiler,[1] Carla Hackworth,[1] Kali Holcomb,[1] Albert Boquet,[2] Elaine Pfleiderer,[1] Douglas Wiegmann,[3]
and Scott Shappell[4]
Federal Aviation Administration[1], Embry-Riddle Aeronautical University[2], Mayo Clinic[3], Clemson University[4]

General aviation (GA) accidents that occurred in Alaska versus the rest of the United States were compared using the Human Factors Analysis and Classification System (HFACS). Overall, categorical differences among unsafe acts (decision errors, skill-based errors, perceptual errors, and violations) committed by pilots involved in accidents in Alaska and those in the rest of the U.S. were minimal. However, a closer inspection of the data revealed notable variations in the specific forms of unsafe acts within the accident record. Specifically, skill-based errors associated with loss of directional control were more likely to occur in Alaska than the rest of the U.S. Likewise, the decision to utilize unsuitable terrain was more likely to occur in Alaska. Additionally, accidents in Alaska were associated with violations concerning VFR into IMC. These data provide valuable information for those government and civilian programs tasked with improving GA safety in Alaska and the rest of the US.

INTRODUCTION

Considerable effort has been expended over the last several decades to improve safety in both military and commercial aviation. Even though many people have died and millions of dollars in assets have been lost, the numbers pale in comparison to those suffered every year within general aviation (GA). For example, according to the National Transportation Safety Board (NTSB), there were 1,741 GA accidents in 2003 that resulted in 629 fatalities (NTSB, 2005). While the numbers may not register with some, when considered within the context of commercial aviation, the losses suffered annually by GA are roughly equivalent to the complete loss of three commercial passenger Boeing 727's.

Why then has GA historically received less attention? Perhaps it has something to do with the fact that flying has become relatively common as literally millions of travelers board commercial aircraft daily to get from place-to-place. Not surprisingly then, when a commercial airliner crashes, it instantly becomes headline news, shaking the confidence of the flying public.

In contrast, GA accidents happen virtually every day yet they receive little attention and seldom appear on the front page of *USA Today*. Perhaps this is because they happen in isolated places, involving only a couple of unfortunate souls at a time. In fact, unless the plane crashed into a school, church, or some other public venue, it is unlikely that anyone outside the local media, government, or those intimately involved with the accident even knew it happened.

Over the last couple of years, GA has deservedly received increasing attention from the FAA (FAA Flight Plan 2004-2008) and other safety professionals. Indeed, several groups from the government (e.g., the FAA's Civil Aerospace Medical Institute; National Institute of Occupational Safety and Health), private sector (e.g., the Medallion Foundation), and universities (e.g., University of Illinois, Johns-Hopkins University) have conducted a number of studies examining GA accident causation.

Alaskan Aviation

It is of note that many of these efforts have focused on Alaska, where aviation is the primary mode of transportation. Alaska is known for its varied and often unique landscape and when this is considered with temperamental weather and seasonal lighting conditions, even the most experienced pilot would have to agree that Alaskan aviation represents some of the most difficult flying in the U.S., if not the world. The combination of factors mentioned above, the number of GA accidents that are occurring in Alaska and the FAA's accident reduction goal (FAA Flight Plan 2004-2008) were factors in our decision to implement this study.

Human Error and General Aviation

A variety of studies have been conducted in an attempt to understand the causes of GA accidents. Most have focused on contextual factors or pilot demographics, rather than the underlying causes of the accidents. When the leading cause of accidents, human error, has been addressed, it is often only to report the percentage of accidents associated with aircrew error in general or to identify those where alcohol or drug use occurred. What is needed is a thorough human error analysis. Previous attempts to do just that have met with limited success (O'Hare, Wiggins, Batt, & Morrison, 1994; Wiegmann & Shappell, 1997). This is primarily because human error is influenced by a variety of factors that are usually not addressed by traditional classification schemes (Shappell & Wiegmann, 1997). Yet, with the development of the Human Factors Analysis and Classification System (HFACS) previously unknown patterns of human error in aviation accidents have been uncovered (Shappell & Wiegmann, 2001; Wiegmann & Shappell, 2001a).

METHOD

GA accident data from calendar years 1990-2002 were obtained from databases maintained by the NTSB and the FAA's National Aviation Safety Data Analysis Center (NASDAC). In total, 24,978 GA accidents were extracted for analysis. Only accidents occurring during 14 CFR Part 91 operations were included (22,987 cases). This analysis was primarily concerned with powered aircraft and thus the data were further restricted to include only accidents involving powered fixed-wing aircraft, helicopters, and gyrocopters. The remaining 22,248 accidents were then examined for aircrew-related causal factors. In the end, 17,808 accidents were included in the database that were associated with some form of human error and submitted to further analyses using the HFACS framework.

RESULTS

When using HFACS to examine the GA accident data, the majority of the accidents are coded with either a precondition for unsafe acts or an unsafe act. This is due primarily to the fact that there is typically not much of an organizational structure or supervisory influence on the majority of GA pilots, as compared to their counterparts conducting commercial or "for hire" operations.

Indeed, with few exceptions (e.g., flight instructors and flight training institutions), the top two tiers of HFACS (unsafe supervision and organizational influences) remained sparsely populated when examining the GA accidents leaving the majority of causal factors within the bottom two tiers of HFACS. Consequently, the balance of this report will focus only on the unsafe acts of the operator level of the HFACS framework.

Unsafe Acts of Operators (Aircrew)

An overall review of the GA accident data yielded the following results (see Figure 1). The most prevalent error noted in the accident data over the past decade was skill-based errors (73%), followed by decision errors (28%), violations (13%), and perceptual errors (7%).[1] The relatively flat lines in the types of unsafe acts across the years suggest that past intervention strategies have had little differential impact on any particular category of error.

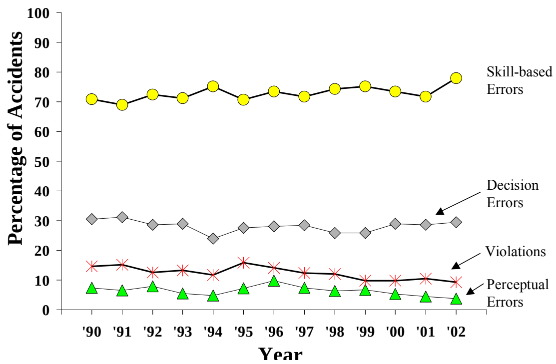


Figure 1. Overall review of general aviation data for HFACS unsafe acts.

To obtain a better sense of how human error differences between Alaska and the rest of the United States (RoUS) are represented in the data, the error types were broken out accordingly (Figure 2). The analysis of the unsafe acts revealed that there were slightly more decision errors, fewer skill-based errors, perceptual errors and violations in Alaska than there were in the RoUS.

Note, the following analyses did not distinguish between those pilots who were native to Alaska and were involved in an accident versus those who were less familiar with the state. That being said, the numbers for Alaska reflect the accidents that occurred within the physical boundaries of the state.

[1] These percentages do not add up to 100 because an accident could be assigned more than one HFACS code (i.e., DE, SBE, PE, etc..).

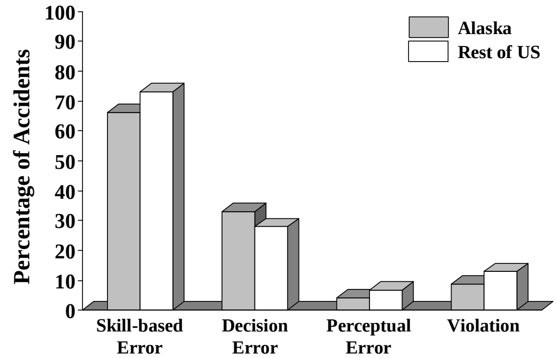


Figure 2. Percentage of accidents associated with each of the unsafe acts of the operator.

Skill-based Errors. Differences that existed between Alaska and the RoUS were fairly consistent across the years of study, with slightly more skill-based errors associated with accidents in the RoUS (see Figure 3). The only exception involved 1991, 1996, and again in 2002 where the percentages were nearly equal.

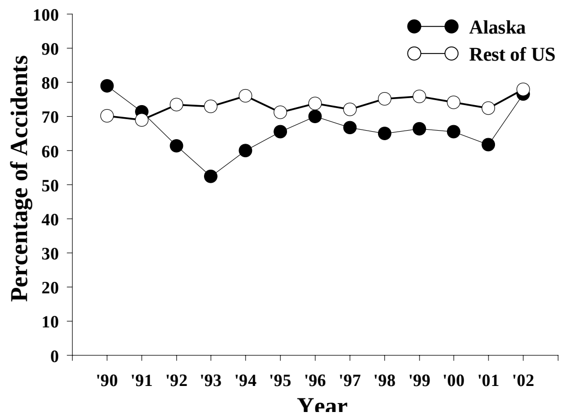


Figure 3. Skill-based errors broken out by Alaska versus the RoUS.

Differences between Alaska and the RoUS were more distinct when the actual type of skill-based error was compared (Table 1). For instance, directional control was the most frequently cited skill-based error for both Alaska (19%) and for the rest of the U.S. (13%). Pilots in Alaska were more likely to experience a loss of directional control of their aircraft than those in the rest of the U.S. (odds ratio = 1.593, X^2 = 33.400, p <.001). Additionally, inadequate compensation for wind conditions was almost three times more likely to occur in Alaska, (odds ratio = 2.884, X^2 = 150.893, p <.001). Conversely, pilots in the rest of the U.S. were almost two times more likely to demonstrate airspeed errors than those in Alaska, (odds ratio = 1.733, X^2 = 20.652, p <.001).

Table 1. Top 5 Skill-based errors occurring for Alaska and the rest of the U.S.

Alaska	N (%)	RoUS	N (%)
Directional Control	206 (18.6%)	Directional Control	2139 (12.6%)
Compensation for Wind Conditions	170 (15.4%)	Airspeed	1932 (11.3%)
Stall	88 (8.0%)	Stall	1312 (7.7%)
Airspeed	76 (6.9%)	Aircraft Control	1310 (7.7%)
Ground Loop/Swerve	50 (4.5%)	Compensation for Wind Conditions	1009 (5.9%)

Decision Errors. To better understand the complexity of the decision errors that were occurring in the accidents for both Alaska and the rest of the U.S., a fine-grained analysis of the data was conducted. Figure 4 illustrates the decision error trends for Alaska and the rest of the U.S. across the thirteen-year period from 1990-2002. With the exception of 1990, 1991, and 2002 any difference that did exist was remarkably consistent across years of the study.

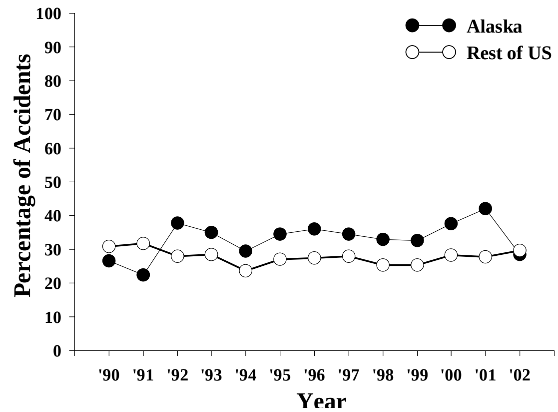


Figure 4. Decision errors broken out by Alaska versus the rest of the U.S.

Upon closer examination, the largest proportion of decision errors in the RoUS involved in-flight planning/decision making, accounting for 19% of those observed. However, the top decision error for pilots flying in Alaska dealt with decisions to utilize unimproved landing, takeoff, taxi areas, or unsuitable terrain. As a matter of fact, those flying in Alaska were almost 15 times more likely to takeoff and land from unsuitable terrain than those in the RoUS (odds ratio = 14.703, X^2 = 829.461, p <.001). A break-out of the top 5 decision errors for Alaska versus the rest of the U.S. is presented in Table 2.

Table 2. Top 5 Decision errors occurring for Alaska and the RoUS.

Alaska	N (%)	RoUS	N (%)
Unsuitable Terrain	193 (40.5%)	In-flight Planning/ Decision	1002 (18.7%)
In-flight Planning/ Decision	59 (12.4)	Planning/ Decision	374 (7.0%)
Aborted Takeoff	28 (5.9%)	Refueling	351 (6.5%)

Planning/ Decision	19 (4.0%)	Remedial Action	339 (6.3%)
Go-around	18 (3.8%)	Go-around	336 (6.3%)

Violations. In general, violations were associated with less than 20% of GA accidents (Figure 5). For the entire U.S. sample, nearly 50% of these accidents resulted in a fatality. When examining accidents in Alaska separately from the RoUS, differences were found. Accidents involving violations in Alaska were 9 times more likely to result in a fatality (odds ratio = 9.248, X^2 = 127.606, p <.001); whereas, those that occurred in the rest of the U.S. were 4 times more likely to result in a fatality, (odds ratio = 4.410, X^2 = 1054.059, p <.001).

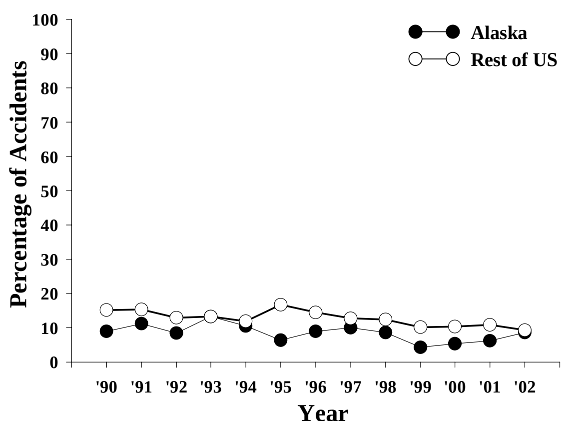


Figure 5. Violations broken out by Alaska versus the RoUS.

A closer look at the types of violations revealed that the most frequently cited violation for all GA accidents was Visual Flight Rules (VFR) flight into instrument meteorological conditions (IMC), (Table 3). VFR flight into IMC alone accounted for one-third of the violations in the Alaska data and was over two and a half times more likely to occur there than in the RoUS (odds ratio = 2.629, X^2 = 22.467, p <.001). Furthermore, when the weather-related violations were combined (VFR into IMC, flight into known adverse weather, and flight into adverse weather), nearly half of the violations in the Alaska data were represented.

Table 3. Top 5 Violations occurring for Alaska and the rest of the U.S.

Alaska	N (%)	RoUS	N (%)
VFR into IMC	38 (32.5%)	VFR into IMC	369 (15.5%)
Aircraft Weight & Balance	13 (11.1%)	Operation with Known Deficiencies	261 (10.9%)
Procedures/ Directives	12 (10.3%)	Procedures/ Directives	248 (10.4%)
Flight into Known Adverse Weather	11 (9.4%)	Flight into Known Adverse Weather	212 (8.9%)
Operation with Known Deficiencies	8 (6.8%)	Aircraft Weight & Balance	149 (6.2%)

DISCUSSION

On the surface, there were no major differences between Alaska and the rest of the U.S. with regard to the overall pattern of human error. If anything, there were slightly more decision errors associated with accidents occurring in Alaska and fewer skill-based errors, perceptual errors, and violations. This information is similar to research in other aviation operations, which identified skill-based errors as the most commonly occurring type of error (Shappell & Wiegmann, 2003; Wiegmann & Shappell, 2001b; 2003).

The accident data suggest that aircraft handling should be taken into account when determining where interventions should be applied. For instance, any training (both *ab initio* and recurrent) along these lines should include control of the aircraft on the ground (e.g., ground loops), crosswind landings, avoiding and recovering from stalls, and general control of the aircraft in flight. Given the inherent risk associated with some of these maneuvers, it makes sense to utilize modern simulators during this training. Unfortunately, it is unclear whether there would be adequate transfer of training for these specific tasks to make simulation training viable. Therefore, before utilizing simulation to address these issues, research needs to be conducted to determine the best role simulators might play. In the meantime however, it appears necessary to emphasize these topics during actual in-flight training.

The only notable exception among the HFACS casual categories involved decision errors. Specifically, pilots in Alaska were more likely to utilize unsuitable terrain for landing, taxi, and takeoff. It would appear that educating aviators on the hazards of utilizing frozen rivers or gravel bars, for example, may reduce these types of errors. However, it may be that there are simply more "improved" areas in the RoUS, providing pilots with more options in case of an emergency (i.e., alternate airports, highways, roads, etc.) in which case education in and of itself may not prove successful. Additionally, it is worth noting that "unsuitable terrain" was a classification imposed by the NTSB investigators after the fact, and the moment-to-moment judgment of how suitable terrain may be during a flight may be influenced by factors not considered fully in post hoc analyses.

Also of concern in both Alaska and the rest of the U.S. was in-flight planning/decision making. After all, decisions made during flight are often more critical than those occurring on the ground. Thus, when confronted with important decisions during flight, pilots are often under pressure to be right the first time while using limited information. Scenario-based training along these lines like that provided within the FAA-Industry Training Standards (FITS) program may improve decision-making in the cockpit, particularly if examples are drawn from the accident record.

Of the unsafe acts that aircrew commit, addressing violations may be the most difficult and complex. Recall that violations are the "willful" disregard for the rules and as such are not necessarily something that can be easily deterred or mitigated. Nevertheless, since nearly half of violations involved fatalities, behaviors like VFR flight into IMC are of great concern to the FAA and other aviation safety professionals.

Even though the percentage of accidents associated with violations did not differ markedly between Alaska and the RoUS, the specific types of violations did differ in meaningful ways. In particular, when intentional VFR flight into IMC and other adverse weather conditions were combined, an alarming 47% of the violations occurring in Alaska were accounted for (27% for the rest of the U.S.). Exactly why a larger proportion was observed in Alaska remains unknown, but one reason may be the rapid weather changes that often occur, especially around mountainous areas.

Current interventions like weather cameras in mountain passes and other locations have proved useful by providing pilots with access to real-time weather information and therefore allowing them to make informed decisions. In addition, the Medallion Foundation has provided GA pilot training using high-resolution flight simulators capable of producing simulated weather and lighting conditions and terrain depictions which are all appropriate to Alaska. With this technology, pilots are able to safely navigate through Alaska and see what flying through places such as Merrill Pass in adverse weather conditions could entail, a difficult task to successfully perform in clear conditions.

Alaska, as perhaps the FAA's largest aviation laboratory, has been the testbed for advanced avionics like those associated with the Capstone project. Enhanced weather radar, global positioning sensors, Automated Dependent Surveillance – Broadcast (ADS-B), and other cutting-edge technologies provide a more accurate picture of how the weather, terrain and traffic situation actually look from inside the cockpit. These technologies have proven useful with 14 CFR Part 135 (commuter) operations (Williams, Yost, Holland, & Tyler, 2002). However, their efficacy within GA remains to be seen.

CONCLUSIONS

In recent years, a growing concern has been directed toward GA accident rates. The FAA Administrator has set a goal of a 20% reduction in GA accidents by fiscal year 2008. If this goal is to be realized, interventions that target the underlying human causes as identified in this analysis need to be developed. Only then can any great strides in improving the GA accident rate be achieved.

REFERENCES

Federal Aviation Administration. (n.d.) FAA Flight Plan 2004-2008. Retrieved August 24, 2004, from http://www.faa.gov/AboutFAA/FlightPlan.cfm.

Federal Aviation Administration. (n.d.) FAA-Industry Training Standards (FITS). Retrieved July 29, 2005, from http://www.faa.gov/education_research/training/fits/

National Transportation Safety Board. (2005). Table 10. Accidents, Fatalities, and Rates, 1985 through 2004, U.S. General Aviation. Retrieved July 12, 2005, from http://www.ntsb.gov/aviation/Table10.htm

O'Hare, D., Wiggins, M., Batt, R. & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37*(11), 1855-1869.

Shappell, S. & Wiegmann, D. (1997). Why would an experienced aviator fly a perfectly good aircraft into the ground? *International Symposium on Aviation Psychology*, The Ohio State University, 26-32.

Shappell, S. & Wiegmann, D. (2001). Beyond Reason: Defining the holes in the Swiss Cheese. *Human Factors in Aviation Safety, 1*(1), 59-86.

Shappell, S. & Wiegmann, D. (2003). *A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990-1998.* (DOT/FAA/AM-03/4). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Wiegmann, D.A. & Shappell, S.A. (1997). Human factors analysis of post-accident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology, 7*(1), 67-81.

Wiegmann, D.A. & Shappell, S.A. (2001a). Human error perspectives in aviation. *The International Journal of Aviation Psychology, 11*, 341-357.

Wiegmann, D.A. & Shappell, S.A. (2001b). A *human error analysis of commercial aviation accidents using the human factors analysis and classification system (HFACS).* (DOT/FAA/AM-01/3). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Wiegmann, D. & Shappell, S. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system.* Ashgate Publishing Company: Aldershot, Great Britain.

Williams, K., Yost, A., Holland, J. & Tyler, R. (2002). *Assessment of advanced cockpit displays for general aviation aircraft – The Capstone Program.* (DOT/FAA/AM-02/21). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Aviation Safety Inspector (ASI) Training for Technically Advanced Aircraft

Carla Hackworth, Ph.D.
Cristina Bates
Kali Holcomb, A. A.
Cristy Detwiler, B. A.
Civil Aerospace Medical Institute
Oklahoma City, OK

Recently, there has been an emergence of technically advanced "glass cockpit" aircraft, within general aviation. Aside from technical challenges presented by the design of these advanced avionics systems, there are difficulties in acquiring a conceptual understanding of the functions offered by the avionics, developing system monitoring skills and habits, developing mode management and awareness skills, understanding when and when not to use automation, and maintaining manual flying skills. Operating aircraft with advanced avionics requires an additional set of knowledge elements and skills. Currently, FAA aviation safety inspectors (ASIs) are required to inspect technically advanced aircraft, check certified flight instructors, and conduct surveillance of designated pilot examiners who are certifying pilots operating technically advanced aircraft. Therefore, the aim of this project is to provide general aviation safety inspectors with the skills needed regarding technically advanced aircraft.

INTRODUCTION

Technically advanced aircraft (TAA) are becoming more prevalent in the General Aviation fleet (AOPA, 2005). Recently, there has been an emergence of technically advanced "glass cockpit"aircraft within general aviation. TAA are generally equipped with IFR GPS navigation equipment with a moving map or a multifunction display (MFD) that displays weather, traffic, or terrain, and has an integrated autopilot (AOPA, 2005; FAA, 2003). Aside from the technical challenges presented by the design of these advanced avionics systems, there are difficulties in acquiring a conceptual understanding of the functions offered by the avionics. Paul Craig and colleagues are examining how best to teach new general aviation pilots in these sophisticated systems (Craig, Bertrand, Dornan, Gossett, & Thorsby, 2005). Casner (2005) generated a detailed list of the knowledge and skills required to be proficient in TAA. Included are the human factors considerations that are required such as the impact of advanced navigation displays on situation awareness and aeronautical decision making given available weather information. These are simply a few of the many factors that are present when considering TAA. Advanced automation introduces both pros and cons; however, regardless of where energy is focused, the recent influx of these aircraft will require additional learning for many in aviation including instructors, examiners, and inspectors.

FAA aviation safety inspectors (ASIs) are required to inspect TAA, check certified flight instructors, and conduct surveillance of designated pilot examiners who are certifying pilots operating TAA. However, many of the ASIs within the FAA workforce completed flight training prior to the entry of advanced avionics.

Therefore, general aviation ASIs need to be more knowledgeable of the capabilities, limitations, and the normal and emergency operating procedures in these aircraft so that they may safely and competently perform their inspection, checking, and surveillance functions for general aviation operators who have these types of aircraft.

AFS-500 has been tasked with developing the "Technically Advanced Aircraft Prerequisite Study Course" and "Qualification for Technically Advanced Aircraft (TAA)" courses for ASIs. The proposed Prerequisite course will provide an overview of three major TAA electronic flight systems used in general aviation. The Qualification course will instruct ASIs how to evaluate pilots and DPEs who operate a TAA.

The aim of this project is to insure that aviation safety inspectors are provided with the skills needed for their job regarding TAA.

.

METHOD

Two courses, the Technically Advanced Aircraft Prerequisite Study course and the Qualification for Technically Advanced Aircraft (TAA) course, have been developed to educate general aviation ASIs on the capabilities, limitations, and the normal and emergency-operating procedures in TAA. AFS-520 provided a grant to Embry Riddle for the development of the Qualification course. During the instructional analysis and design of the course, it was determined that a prerequisite study course would also be needed to prepare inspectors for the material to be presented in the Qualification course.

Course Descriptions

The initial prerequisite course is to provide ASIs with an overview of three major TAA electronic flight systems used in general aviation. The evaluation course will instruct ASIs how to evaluate pilots and DPEs who operate a TAA. The course will provide ASIs with the minimal proficiency standards required to operate a TAA.

After completing each of these courses, ASIs will complete course evaluations on their impressions of the course content and the extent to which the courses prepared them to perform their TAA job functions. Additionally, inspectors are evaluated once at the completion of the Prerequisite course and twice in the Qualification course. The Prerequisite course evaluation is an end-of-course open-book test. The two evaluations in the Qualification course are: a single engine airplane "glass cockpit" check, and an end-of-course check of evaluation skills.

Participants will be Aviation Safety Inspectors that enroll in FAA-sponsored courses (18803 and 18830) organized through Embry Riddle Aeronautical University. In the future, additional university campuses will assist in the delivery of these courses.

Participants will receive a prerequisite course evaluation and a qualification course feedback survey. The surveys address ASIs perceptions of their proficiency as a result of the courses and course content.

Prerequisite Course Survey Content

Respondents will rate the degree to which the course material was related to their job duties, how well they can explain symbols used for navigation and terrain on the multifunction display and how prepared they are to perform system failures in TAA.

Evaluation Course Content

Respondents will rate how effective the course material was in preparing them for surveillance of TAA, how well the check-ride allowed them to demonstrate their proficiency, how well they understand the human factors implications within TAA, and the extent of their understanding of simulating TAA system failures. Additionally, participants will be asked if they have had any previous hands-on experience with TAA.

Participants will be assured that the surveys are voluntary and that they may choose not to answer any particular question. The right to refuse to participate will be inherent in the survey process, as participants will only complete the survey if they choose to do so.

RESULTS

The first collection of evaluation data is set to commence in October 2005 at ERAU in Daytona Beach. Data will be analyzed using simple descriptive statistics (e.g., frequencies, means, and proportions). Open-ended questions and comments will be content-coded and analyzed with descriptive statistics as well. Summary reports will be created by course provided the courses have at least 8 respondents. We plan to use the data to improve the courses and ensure that ASIs are learning the skills required to perform their duties regarding TAA.

REFRERENCES

AOPA ASF (2005). Technically advanced aircraft: Safety and training. http://www.aopa.org/asf/publications/taa_1_6.pdf.

Casner, S. (2005; in preparation). Technically advanced aircraft flying handbook: Extended table of contents.

Craig, P., Bertrand, J.E., Dornan, W., Gossett, S., & Thorsby, K.K. (2005). Ab initio training in the glass cockpit era: new technology meets new pilots. *Proceedings of the 13th International Symposium of Aviation Psychology*, Oklahoma City, Oklahoma. April 2005.

Federal Aviation Administration (2003). General Aviation TAA Safety Study.

Developing a Methodology for Assessing Safety Programs Targeting Human Error in Aviation

Scott Shappell, Ph.D.
Clemson University
Clemson, SC

Douglas Wiegmann, Ph.D.
Mayo Clinic
Rochester, MN

There is a need to develop an effective methodology for generating comprehensive intervention strategies that map current and proposed safety programs onto well-established forms of human error. Two separate studies were conducted using recommendations from NTSB accident investigations and several joint FAA and industry working groups. The goal was to validate a proposed framework for developing and examining safety initiatives targeting human error in aviation. The results suggest five approaches to reducing human factors associated with aviation accidents. When combined with the Human Factors Analysis and Classification System (HFACS) the resulting Human Factors Intervention Matrix (HFIX) will provide a useful tool for evaluating current and proposed aviation safety programs.

INTRODUCTION

The NTSB, FAA, and other safety organizations have committed extraordinary resources to prevent civilian aviation accidents. As a result, aviation in the U.S., particularly commercial aviation, has become one of the safest modes of transportation. Still, accidents can happen, often repeating the same sequence of events played out many times before. As a result, we are left with the regrettable truth that there are really very few "new" accidents, just different players.

So if there really are few "new" accidents, why has the aviation accident rate remained relatively stable over the last several years? Perhaps it has something to do with the current state of aviation safety. Truth be told, the industry is extremely safe and the easy fixes have been identified and remedied. What remains to be addressed is that small fraction of accidents attributable to perhaps the most complex problem facing aviation today – human error.

Unfortunately, while previous safety programs may have impacted other areas of aviation, there has been little evidence that they have had a significant impact on any specific type of human error. That is to say, the percentage of accidents associated with aircrew error (e.g., skill-based errors, decision errors, perceptual errors, and violations) has remained relatively stable since 1990 (Wiegmann & Shappell, 2003).

What this implies is that intervention strategies implemented in the 1990's had at best ubiquitous effects on the errors and violations committed by aircrew. More likely however, there has been no sustained impact of any particular intervention program (Shappell, Detwiler, Holcomb, Hackworth, Boquet, & Wiegmann, in review). The latter should come as no surprise given that few studies of aircrew and supervisory error had been conducted using a human factors approach to accident causation. Furthermore, there has been no systematic human factors examination of the current or proposed safety programs aimed at addressing human error. For that matter a "human factors" analysis of safety programs even possible?

At least one study (Wiegmann & Rantanen, 2003) suggests that such an analysis can be performed using a set of standards derived from the same body of literature used to develop HFACS. In their book Wiegmann and Shappell (2003) describe an intervention taxonomy clustered around four broad categories:

1. Environment (control of temperature, noise, vibration, lighting, etc)
2. Human (personnel selection, incentives, training, teamwork, communication, etc.)
3. Machine (engineering design, capacity, etc.)

4. Task (ordering/timing of events, procedures, standardization, etc.)

Using this framework, Wiegmann and Rantanen (2003) examined a variety of technologies developed by NASA's aviation safety program. From energy absorbing seats, restraints, and structures to synthetic vision, each safety program was classified within one of the four intervention categories. Their initial classification revealed that NASA's primary intervention strategies targeted the machine rather than the human, environment, or task. Furthermore, when they examined NASA technologies using the HFACS framework, it was determined that nearly half of the technologies that NASA was developing were rated as having no impact on aircrew error. Those that did primarily targeted decision errors, by providing better information, automation, and training. An even smaller percentage of the technologies were aimed at aircrew error in general and only one of the products primarily targeted skill-based errors – the number one problem facing both commercial and general aviation.

Purpose

This report describes two studies that build upon the methodology originally described by Wiegmann and Shappell (2003) and used by Wiegmann and Rantanen (2003) with NASA safety programs. The first study describes an independent validation of the four intervention methodologies using safety recommendations from the NTSB. The second describes the examination of proposed FAA aviation safety programs using a prototype intervention matrix that maps the unsafe acts of operators (i.e., skill-based errors, decision errors, perceptual errors, and violations) onto several intervention approaches.

STUDY 1:
ANALYSIS OF NTSB RECOMMENDATIONS

Investigating accidents, identifying potential interventions, and issuing safety recommendations are central to any safety program. Ideally, safety recommendations, when adopted by organizations, will positively influence future operations in the field and thereby improve overall system safety. However, recommendations are just that … *recommendations* and as such are not always adopted. Moreover, they are often based solely on isolated events or at best a few events over a very short period of time rather than more global analyses of the system as a whole. While these interventions may solve a local or single-point problem, they often do not have far-reaching impact.

Further complicating matters, many domains like aviation and their corresponding safety boards have traditionally strong relations with quantitative disciplines like engineering and physics. While these organizations may be especially adept at dealing with mechanical issues they tend to be less robust when dealing with organizational or human-centered aspects of accidents such as human error, organizational failure, communication, and risk assessment (Stoop, 2003).

Recognizing this, the NTSB, like many safety entities has integrated human factors experts into their organization presumably leading to recommendations that address the entire system rather than a single engineering or mechanical aspect, per se. However, employing human factors experts alone does not necessarily translate into a breadth of interventions. A reasonable question to ask then is what specific intervention approaches does the NTSB employ? In other words, does the NTSB tend to be uni-dimensional (like NASA) or multi-dimensional with regard to specific intervention approaches?

Method

NTSB Safety Recommendations

To examine this question, aviation safety recommendations associated with commercial (14 CFR Part 121 – air carrier and Part 135 – commuter) aviation accidents occurring between 1998 and 2004 were obtained from the NTSB database. Of the 147 commercial aviation accidents reports that were completed at the time of this study, 622 unique safety recommendations were identified. However, several of the recommendations consisted of compound solutions. In those cases, the original recommendation was separated into sub-recommendations yielding a revised list of 872 unique recommendations for further analysis.

Clustering Process

The recommendations were then independently clustered into categories by two based on their similarities. The analysts were not instructed to use any predefined taxonomy or classification scheme. They were simply instructed to independently assign each recommendation to categories of their choosing based upon the nature of the recommendation.

Not surprising given the vagueness of the instructions, there were some differences in the terms used by the two analysts but there were also strong similarities. Wherever disagreements occurred, the analysts were asked to discuss their clustering heuristic and to agree on a single classification scheme. In the end, all 872 recommendations were classified based on their underlying similarities by two independent analysts, who later came to a consensus on the number and labels for each of these clusters.

Results

Ultimately, the analysts generated nine unique categories of recommendations, which included the design of parts/displays, procedures, communication, training, requests to conduct focused studies, rules, manuals, inspection, and human resources. These nine categories were then further grouped into four larger categories based on their similarities: 1) administrative/organizational; 2) mechanical/ engineering; 3) human/crew; and 4) task/mission. Each category and their accompanying subcategories are briefly described in Table 1.

Distribution of recommendations

From a global perspective, it appears that roughly two-thirds of the recommendations were administrative/ organizational or mechanical/ engineering fixes while nearly a quarter of the recommendations were aimed at either the task or mission. Surprisingly few interventions directly targeted operators (aircrew) even though previous studies repeatedly show that more major accidents have been attributed to human error than to any other single cause (Wiegmann & Shappell 2003, Boquet, et.al., in review; Detwiler, et.al., in review; Shappell, et.al., in review). It has also been observed that wider systemic issues, including the managerial and regulatory context of aviation operations, were also mentioned in a large number of reports (Holloway & Johnson, 2004; Johnson, in review), even though this does not appear to be reflected in the accident record.

A closer examination revealed that similar to Wiegmann and Ratannen's study of NASA safety programs, design fixes constituted the largest percentage of any individual type of recommendation made by the NTSB (23.17%) - nearly twice as many as any other category. Considerably fewer recommendations were aimed at procedures, training, information management/ communication, and the other subcategories.

Summary

When examining the breadth and scope of NTSB recommendations even at this level, it appears that current aviation safety recommendations tend to focus more on improving the design of systems or some manner of organizational change rather than focusing on the individuals in the field. While these recommendations are obviously well-intentioned and often specific to a particular accident, they may be misplaced or narrow in scope. This may help explain why the percentage of accidents associated with human error has not changed over the last 15 years (Wiegmann & Shappell, 2003; Wiegmann, et. al, in press; Shappell, et. al., in review).

Table 1. Proposed categories and sub-categories of NTSB recommendations.

Administrative/Organizational
Rules/Regulations/Policies: Issuing, modifying, establishing, amending, and/or reviewing policies, rules, or regulations.
Information Management/ Communication: Improvements in disseminating, storing, archiving and publishing information. Also included are recommendations regarding collection of data, issuing information bulletins, advisory circular and reporting activity.
Research/ Special Study: Conducting research to determine the impact of recent technological advances or call for special studies to review processes, develop/validate methodologies, evaluate the feasibility of safety equipment, and/or conduct surveys.
Human Resource Management: Adequacy of staff in specific situations, the need for additional personnel, and the evaluation of individual skills of employees.
Mechanical/Engineering
Design/Repair: Specific manufacturing changes including the design of parts. Also included is the modification, replacement, removal and/or installation or repair of parts and equipment.
Inspection: Maintenance inspections, overhauling, detecting damage including day-to-day operations such as inspecting fuel, oil level, and recommended safety checks.
Human / Crew
Training: Reviewing, developing, and implementing training programs. Also included is the training of personnel in handling emergencies.
Task/Mission
Procedures: Amending, reviewing, modifying, revising, establishing, developing, and validating procedures.
Manuals: Reviewing, revising, issuing, amending, and modifying manuals, bulletins, checklists, and other instructions or guidance.

The findings of Study 1 suggest that there are at least four broad categories of interventions that appear tenable within the aviation industry: Administrative/ Organizational, Human/Crew, Mechanical/ Engineering, and Task/Procedure. These four approaches differed slightly from those previously proposed by Wiegmann and Shappell (2003) and utilized by Wiegmann and Rantanen (2003) to analyze NASA safety programs. One category that naturally surfaced from the present analysis, but was missing from the Wiegmann and Rantanen study, was Administrative/Organizational interventions. In contrast, “environmental” interventions did not appear in the current study but were present in the NASA study (Wiegmann & Rantanen, 2003).

In the end, the question is not whether or not there are three, four, five, or more approaches to identify potential accident interventions as much as there is definitively more than one. Exactly what those approaches are remains to be fully explored. However, the five approaches identified between the present study and the investigation conducted by Wiegmann and Rantanen (2003) is a reasonable first start.

STUDY 2
HFIX ANALYSIS OF JSAT/JSIT RECOMMENDATIONS

Identifying viable approaches for intervening is only the first step. The ability to map interventions onto specific types of human error is also important. In other words, simply generating a variety of interventions across several domains, whether they are human, mechanical, environmental, etc., is likely to be ineffective unless such interventions directly target the problem area.

Given that human error continues to be the largest contributor to commercial and general aviation accidents, it makes sense to map different interventions against specific error forms. What is needed is a theoretical framework that captures the underlying causal mechanisms of human error that align with the intervention approaches identified in Study 1.

Such an error framework already exists and is widely used within the aviation industry. This framework, the Human Factors Analysis and Classification System (HFACS) describes two general categories of unsafe acts that operators commit: *errors* – the honest mistakes individuals make every day, and *violations* – the willful disregard for the rules and regulations of safety[1]. Within those two overarching categories, HFACS describes three types of errors (decision, skill-based, and perceptual) and two types of violations (routine and exceptional). Each has been described extensively in previous reports (e.g., Wiegmann & Shappell, 2003).

Human Factors Intervention Matrix (HFIX)

A prototype matrix, called the Human Factors Intervention Matrix (HFIX), pits the unsafe acts individuals commit against the five different intervention approaches presented above (Figure 3). The utility of such a framework seems intuitive. For example, if one were interested in developing interventions to address decision errors, the goal would be to identify prospective interventions within each approach (i.e., organizational/administrative, human/crew, etc.), thereby ensuring that the widest array of interventions were considered. By mapping prospective interventions onto the matrix it would be readily apparent if the scope of a proposed program was uni- or multi-dimensional.

[1] A complete description of the entire HFACS framework including all 4 tiers and 19 causal categories can be found in Wiegmann and Shappell, 2003.

	Organizational/ Administrative	Human/ Crew	Technology/ Engineering	Task/ Mission	Operational/ Physical Environment
Decision Errors					
Skill-based Errors					
Perceptual Errors					
Violations					

Figure 3. The "Human Factors Intervention matriX" (HFIX).

Alternatively, a framework like HFIX could be used proactively to determine which areas an organization has "covered" and where gaps exist in the current safety program given current trends in the error data. For instance, if you knew that the largest threat to safety within your organization was skill-based errors, followed by decision errors, violations, and perceptual errors (as is the case with general and commercial aviation in the U.S.), HFIX could be used to determine if your proposed and future interventions have the potential to address those needs and which areas are currently being targeted.

Hence, the purpose of Study 2 was to determine if such an approach could be used within the FAA and which types of human error might be affected by current and future interventions. In a sense, this analysis would provide a "benchmark" of current FAA intervention efforts.

FAA *Safer Skies* Initiative

As part of the FAA's *Safer Skies* initiative, several Joint Safety Analysis Teams (JSATs) and Joint Safety Intervention Teams (JSITs) were formed from experts in the government, private sector, industry, and academia to address civilian aviation accidents. Particularly germane to this study were outcomes derived from the JSAT and JSIT teams examining accidents associated with:

- Controlled flight into terrain
- Approach and landing

- Loss of control
- Runway incursions
- Weather
- Pilot decision making

Method

JSAT and JSIT recommendations

Final reports from the selected JSAT and JSITs were collected by researchers at the Civil Aerospace Medical Institute (CAMI). After eliminating duplicate recommendations, a comprehensive list of recommendations was compiled electronically for classification. The final list of 614 unique recommendations was then randomized to reduce bias.

Categorization of the Data

Eighteen Master of Aeronautic Science candidates were recruited from Embry-Riddle Aeronautical University to classify the recommendations. Each had experience in the aviation community as either a pilot, maintainer, or at an administrative level and all had successfully completed a minimum of one graduate level human factors course.

After a roughly 4-hr training session on the HFACS and HFIX frameworks, participants were randomly assigned to one of six groups. Each 3-person team was then randomly assigned roughly 1/6th of the recommendations to classify. Each team member was instructed to independently classify each recommendation into only one of the five intervention approaches (i.e., organizational/administrative, human/crew, mechanical/ engineering, task/mission, or physical environment). In addition, they were instructed to identify any HFACS Unsafe Acts categories they felt the intervention would impact.

After the initial rating, team members were permitted to discuss their classification within their group to resolve any differences. A final, consensus, classification for each recommendation was then provided for further analysis.

Results

The results of both classification tasks are presented in Figure 4. Several observations can be made from the data. First, as with the NTSB recommendations large percentages (36.6%) of JSAT/JSIT recommendations were directed at organizational/administrative levels. Likewise, several (22.2%) of the recommendations involved technological/engineering approaches. However, unlike the NTSB where relatively few recommendations targeted the human, nearly 1/3 of those obtained from the JSAT/JSITs did so.

HFACS UNSAFE ACTS \ INTERVENTIONS	Organizational/ Administrative	Human/ Crew	Technology/ Engineering	Task/ Mission	Operational/ Physical Environmen	
Decision Errors	25.9%	26.9%	13.4%	5.7%	0.8%	72.6%
Skill-based Errors	13.8%	20.7%	12.5%	2.4%	0.2%	49.7%
Perceptual Errors	9.0%	12.7%	12.4%	2.4%	1.1%	37.6%
Violations	14.3%	7.8%	2.8%	1.3%	0.7%	26.9%
	36.6%	32.6%	22.2%	7.3%	1.3%	

Figure 4. Percentage of JSAT/JSIT recommendations classified by intervention approach and specific HFACS unsafe act addressed.

When examining the HFACS classifications, remember that unlike the specific approaches to accident interventions where subjects were instructed to select only one approach, they were permitted to select all of the HFACS Unsafe Act categories that they felt would be impacted by a given recommendation. Therefore, unlike the intervention approaches whose percentages added up to 100%, the total percentages associated with each Unsafe Act category did not.

Perhaps not unexpected, interventions aimed at decision errors were associated with nearly three out of every four JSAT/JSIT recommendations examined. In contrast, skill-based errors were associated with roughly 50% of the recommendations followed by perceptual errors (37.6%) and violations (26.9%). Of note, these numbers are slightly different than the percentage of accidents associated with each type of error where skill-based errors account for between 45-80% of the accidents depending on whether one is talking about commercial or general aviation respectively. Likewise, roughly 1/3 of the accidents were associated with decision errors yet 72.6% of the interventions have some component that will potentially affect pilot decision making.

This is not to say that there should be a one-to-one relationship between the percentage of accidents associated with a given error category and the percentage of recommendations aimed at addressing these errors. After all, it may take more effort to address one error form than another, or more interventions may naturally address pilot decision-making. In either case, the global analysis presented here suggests that additional review of this apparent incongruity is necessary.

Perhaps more important however, was the mapping of each intervention within both the intervention approach and the HFACS Unsafe Acts category (Figure 4). As can be seen (white boxes), three of the 20 possible boxes (organizational/ administrative by decision error, human/crew by decision error, and human/crew by skill-based error) contained 20% or more of the JSAT/JSIT interventions. On the surface this appears to reflect a narrow rather that a broad approach to accident intervention/mitigation by these committees. Not that the interventions contained within these categories will not be effective, just that other, potentially equally viable, interventions may have been overlooked.

It is interesting to note however, that if one examines those boxes that contained between 10-20% of the possible interventions, nearly all of the remaining boxes among the organizational/ administrative, human/crew, and technology/ engineering approaches were included. What were not accounted for were human/crew and technology/engineering approaches dealing with violations of the rules and regulations. Obviously, these approaches might prove beneficial if an organization wanted to modify or curtail a particular unsafe pattern of behavior (e.g., flight into instrument conditions while on a visual flight rules flight plan) through training or technological means.

More notable was the general lack of interventions targeting the specific task/mission of the aircrews or the environment they are faced with. Perhaps a closer examination of the operations these aircrews are engaged in or the environments they are expected to operate in is warranted. In any event, there may have been options that were not considered by these select committees along these lines.

Summary

The results from Study 2 using JSAT/JSIT interventions, although clearly more multi-dimensional than NASA's safety programs, still did not appear to fully address the current accident trends in commercial and general aviation. At least on the surface, it appears that there are weaknesses in the safety program that should be addressed.

For example, there was an apparent bias toward interventions aimed at pilot decision making, particularly those utilizing organizational and human approaches. While this is not inherently bad, previous HFACS analyses suggest that additional effort should be placed on skill-based errors and violations, two areas that appear underrepresented given current trends in the accident data.

Also noteworthy, few interventions attempted to modify/change the task itself or the environment. A closer examination of the actual types of errors may suggest changes in routes people fly or the actual type of flights being flown.

CONCLUSIONS

While HFIX may prove useful when generating comprehensive intervention strategies, organizations simply cannot implement every recommendation. Other factors may need to be considered before employing a given intervention. Factors such as *effectiveness* (i.e., what is the likelihood that it will work?), *cost* (i.e., Can the organization afford the intervention?), *feasibility* (i.e., how easy will it be to implement the intervention or does it actually exist?), and *acceptability* (i.e., will the workforce accept the proposed intervention?) all must be considered.

As such, HFIX may actually be $HFIX^3$ mapping human error against the intervention approaches and evaluation criteria (Figure 5). Although it may appear complex, in reality organizational decision makers utilize this third dimension all the time. However, even without this third dimension, the mapping of specific interventions onto a matrix that combines the five intervention approaches with general categories of human error can provide a broader perspective of the FAA's safety programs.

In sum, safety recommendations are not simply based on empirical findings surrounding an accident. Rather, they are based on one's philosophical view of what actually constitutes a "cause" of an event, coupled with one's own biased view of how changes in human or system behavior can even be accomplished. Therefore, thinking "outside the box" when it comes to generating intervention strategies is extremely difficult to do; yet failure to do so can leave other potentially viable and effective alternatives unexplored. Ideally, the HFIX framework will help safety professionals do just that.

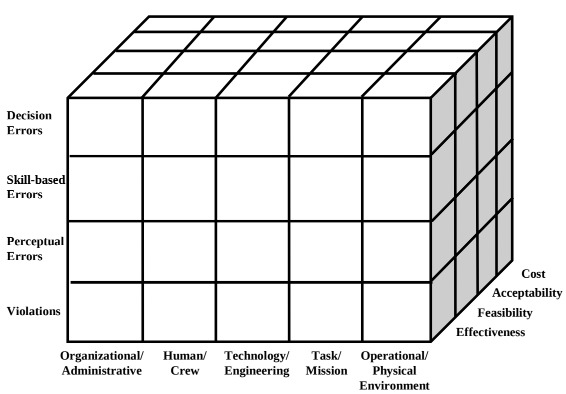


Figure 5. The $HFIX^3$ framework.

REFERENCES

Boquet, A., Detwiler, C., Holcomb, K., Hackworth, C., Wiegmann, D., and Shappell, S. (in review). Human error associated with air medical transport accidents in the United States. Office of Aerospace Medicine Technical Report. Civil Aerospace Medical Institute, Oklahoma City, OK 73125.

Detwiler, C., Hackworth, C., Holcomb, K., Boquet, A., Pfleiderer, E., Wiegmann, D., and Shappell, S. (in review). Beneath the tip of the iceberg: A human factors analysis of general aviation accidents in Alaska verses the rest of the United States. Office of Aerospace Medicine Technical Report. Civil Aerospace Medical Institute, Oklahoma City, OK 73125.

Holloway, C., and Johnson, C. (2004). Distribution of Causes in Selected U.S. Aviation Accident Reports Between 1996 and 2003. *22nd International System Safety Conference, Providence, Rhode Island, Aug2-6. 2004.*

Johnson, C. (in review). On the over emphasis of human error as a cause of aviation accidents: systemic failures and human error in US NTSB and Canadian TSB aviation reports 1996-2003. Submitted to *Ergonomics.*

Reason, J. (1990). *Human Error*. New York: Cambridge University Press.

Shappell, S. and Wiegmann, D. (2003). Reshaping the way we look at general aviation accidents using the human factors analysis and classification system. *Proceedings of the Twelfth International Symposium on Aviation Psychology*, Dayton, Ohio.

Shappell, S. and Wiegmann, D. (2004). HFACS Analysis of Military and Civilian Aviation Accidents: A North American Comparison. *Proceedings of the Annual Meeting of the International Society of Air Safety Investigators*, Gold Coast, Australia.

Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., and Wiegmann, D. (in review). Human error and commercial aviation accidents: A comprehensive, fine-grained analysis using HFACS. Office of Aerospace Medicine Technical Report. Civil

Aerospace Medical Institute, Oklahoma City, OK 73125.

Stoop, J. (2002). Accident investigations: Trends, Paradoxes and opportunities. *International Journal of Emergency management.* 1(2), *170-182.*

Wiegmann, D., Boquet, A., Detwiler, C., Holcomb, K., Faaborg, T., and Shappell, S. (in press). Human error and general aviation accidents: A comprehensive, fine-grained analysis using HFACS. Office of Aerospace Medicine Technical Report. Civil Aerospace Medical Institute, Oklahoma City, OK 73125.

Wiegmann, D., and Rantanen, E. (2003). Defining the Relationship Between Human Error Classes and Technology Intervention Strategies. *Final Technical Report AHFD-03-15/NASA-03-5.*

Wiegmann, D. and Shappell, S. (2001). Human error analysis of commercial aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Aviation, Space and Environmental Medicine*, *72*, 1006-1016.

Wiegmann, D. and Shappell, S. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system.* Burlington, VT: Ashgate.

ENHANCED GENERAL AVIATION DECISION MAKING: WEATHER ASSESSMENT DURING PREFLIGHT PLANNING

Bettina L. Beard, Ph.D. & Colleen T. Geven, ATP, CFI

This report discusses strategies that may be used by pilots to enhance weather decisions during general aviation preflight planning. Current classroom decision training is primarily based on outdated theory as well as anecdotal evidence. More recent research completed in field settings hold important implications for improved training. Here we blend information from current research and theory in human learning, memory, decision making and expertise and apply it to database information, expert opinion, and a look at what is lacking in current training curriculums.

INTRODUCTION

In ground training the pilot learns about aeronautical theories and concepts such as aircraft systems, basic meteorology, weather data interpretation, etc. This basic knowledge is critical toward an ability to move to the next step, which is to apply these concepts and practice basic flying skills. These basic flying skills must be practiced repetitively1 in order to acquire the proficiency necessary to safely control the aircraft. This repetition does lead to learning, however, this learning is not geared toward handling the complexities, the fluidness, of decisions required to expertly pilot an aircraft. Typical training can be inadequate towards these higher-level decisions that will be encountered during all phases of flight. Flying is done in often uncertain, dynamic situations that can include a range of meteorological phenomenon. Pre-flight and in-flight decisions about a course of action are made based on a complex integration of the pilot's experience, weather information from a range of sources, the need to reach a destination, time pressure, personal attitude, human psychological tendencies, etc.

The purpose of this paper is to introduce strategies that will be integrated into aeronautical decision-making training products that are based on current theory and published empirical research. In this report we focus on critical decisions made during the preflight phase.

METHODS

Safety risk identification

We will present a prototype-training product in September 2005. So that this product provides maximum impact toward aviation safety we are constraining the training to high-risk issues. Three methods were used to identify these top issues.

(1) For six months, we participated in bi-weekly discussions with general aviation weather decision domain experts. The results of these discussions confirmed the data obtained from the second method. (2) Lists of 83 search terms (e.g., "VFR flight into IMC", "decision-making,", etc.) were used to conduct a search of the ASRS database. The search was conducted for reports filed by general aviation (GA) pilots between January 1995 to January 2005. Over 500 reports were identified and 68 reports were found to be relevant to this project. (3) Data published in the 2004 Nall Report, which summarizes NTSB accidents, also support our ASRS conclusions.

Please address correspondence to Dr. Tina Beard at NASA Ames Research Center, Moffett Field, CA 94035; Tina.Beard@NASA.gov. This research was supported by funds from FAA AAR-100 and the NASA Aviation Safety and Security Program to Dr. Beard.

The result of applying recent research and theory to the focused issues identified using these three methods will now be addressed in the context of preflight planning.

RESULTS

Assessment of the 68 ASRS reports revealed three prominent findings: poor weather assessment, overestimating piloting capabilities and, distractions leading to aircraft upset (this last finding relates to poor instrument scan when the pilot is distracted by such things as struggling with spatial disorientation, unexpected poor aircraft performance, or difficulty with GPS programming). Here our focus will be on poor weather assessment, which was identified as a major recurring issue arising in 34 of the 68 ASRS general aviation reports.

Specifically pilots either did not obtain adequate weather information and were taken off guard, they did not adequately interpret the weather information, or they underestimated the danger of entering a cloud (because this is an en route topic, it will be discussed in a later publication).

During preflight pilots sometimes choose not to call a Flight Service weather briefer. There appears to be a number of reasons as to why this happens. For example, the pilot has get-there-it-is and rushes through the preflight. Other pilots find it unnecessary to call altogether because the weather appears acceptable just by looking outside or flight route familiarity. Sometimes pilots simply forget to call flight service do to distractions during preflight such as peer pressure or aircraft rental implications. Pilots sometimes decide to rely on automated weather reporting systems instead, such as AWOS for cloud/ceiling and visibility information during their preflight planning, or choose to depart when VFR is not recommended by flight service. Because Flight Service typically errs on the conservative side some pilots may disregard their advisory. This conservatism indicates that current or forecasted weather conditions may be at or below VFR weather minimums and the pilot should investigate why this advisory was provided.

DISCUSSION

We propose that if an enhanced understanding of the weather situation had been obtained during preflight, many of these incidents could have been avoided.

Strategies that may be used to better assess the weather are to

- Learn how to recognize typicality, and therefore anomalies,
- Mentally simulate a course of action,
- Prioritize cues,
- Develop expectancies.

We will now discuss how to implement these strategies.

When first introduced to piloting, novices improve performance for a period of time until reaching an average level of performance, at which point performance remains relatively stable. The striking difference between expert and average performance seems to result not just from how long the pilot has had his piloting license or even how many hours he has logged, but on the quality of the training received, deliberate practice in simulated environments, the particular types of flying experiences that he has encountered, and whether he considers piloting fun.[2]

Novice and expert pilots fly in different worlds. The novice is involved in the moment, such as being distracted with trimming the aircraft, whereas the expert pilot controlling the aircraft is routine and therefore able to constantly assess the overall situation, such as if the actual weather agrees with the forecasted weather. If the novice or intermediate pilot can master a broader, holistic, view of flying, then they may better recognize potential problems and find solutions before meeting a hazardous condition head-on.

Taking a broader view starts with a well-laid flight plan. Developing a detailed flight plan helps the pilot to get mentally in 'the game' by providing him with the big picture. To develop a flight plan the pilot must pull together different resources which forces him to think broadly, to become more familiar with every aspect of the flight leading to more informed in-flight decisions, and to better manage his workload during flight.

A large part of preflight planning includes gathering and interpreting weather information. What sources of weather information should a pilot use for preflight planning? While weather reports on the television are an easy way to keep an eye on the big picture, they cater to weather relevant to people on the ground. Preflight planning should include the use of Internet aviation weather services; such as the prognostic charts, satellite images, text reports and forecasts. The combination of this information from these sources will provide a clearer understanding of an extensive weather product such as an Area Forecast (FA). In a current multi-agency effort entitled the Next Generation Air Transportation System (NGATS) the plan is to develop automated capabilities to integrate weather information for the pilot. However, until a considerably deeper understanding of how to automate decisions in uncertain, dynamic environments is attained, the pilot must continue to be the authority on whether he should continue with his flight.

Because there are numerous sources of weather information that present the data in different formats it is imperative to know what information from which weather product is relevant and how to apply it to the current flight. Expert pilots can extract the relevant information to make a rapid [3-7], acceptable (rather than best) [8-10] decisions. How can novices possibly learn this experience-based ability? To answer this question, we will examine common difficulties pilots face in interpreting the weather.

Meteorological conditions are complex and constantly changing. When a pilot is planning a flight and the weather information is ambiguous, a change of plan may be difficult to justify. This can result in a common error made in piloting called a "plan continuation error"[11]. One strategy used by expert pilots that may reduce the ambiguity of the weather information is to take the time to address any uncertainties about that particular, unique flight. The expert pilot will call back the Flight Service Station (FSS) to ask questions that clarify the ambiguity. ASRS reports show that less experienced pilots may choose not to be assertive because it makes them feel uncomfortable. However, if an expert pilot still does not understand the weather picture, they dig deeper sometimes going to the source - the National Weather Service (NWS). You can get the NWS phone number from FSS or in your Airport/Facility Directory (AFD).

Pictures are much better remembered than words.[12] A pilot may feel bombarded by weather information, for example, the Flight Service Briefer may rapidly speak an Area Forecast (FA). To process and remember this information particularly if the pilot does not have access to a computer, he may draw the FA either on the sectional chart or on a photocopy of an area map (there are charts in the AIM (Aeronautical Information Manual), such as the geographical area designation map in which FA's are often related to). If information about the direction and speed that a system is moving is also drawn onto the map it may help with potential decisions related to choosing an alternate airport.

At times a pilot may be aware of cues in a weather report that suggest the flight should be postponed, but they choose to take-off nonetheless. There are at least three identified reasons for this risky behavior. First, the pilot may have underestimated the level of risk. [13] Here it is suggested that the level of risk may be more clearly seen by imagining, while still on the ground, different scenarios of how the weather may progress, or change, making it less likely that the pilot will be surprised.

A second identified reason that pilots might fly into bad weather is that the pilot may lack the relevant experience and therefore do not recognize a cue as a risk. [14] One strategy that may hasten experience level is for the pilot to observe the weather everyday, all day, even if he is not flying that day. Whenever the pilot looks out the window, or when he goes outside, he should take a look at the clouds. What are they doing today? Why are the clouds shaped as they are? Why is their altitude changing? Finding answers to these questions is good practice, because when it comes time to fly, one of the skills that every pilot should have is to be able to 'read' the clouds. Cloud shape, color and thickness, and altitude can be used as weather indicators. As this skill improves the pilot will begin to correlate the temperature, dew point, humidity, and time of day to the types of clouds that have formed. Also, the expert pilot takes notice of the wind and imagines the conditions visually in his mind. For example, he visualizes how the wind wraps around the tree or whips around the corner of a building. This technique can be helpful during take-off or landing – an unexpected gust of wind won't be so unexpected. By honing in on these skills the novice pilot will have a better understanding of the big weather picture and the cues to watch for before he walks into the Fixed Base Operator (FBO) and calls Flight Service.

A third reason that pilots sometimes fly into bad weather is that they may be committing what is called "frequency gambling" [15-17] which refers to expecting to succeed using a behavior that previously succeeded in a similar risky situation. Even if a pilot simply observes another pilot succeed can lead him into frequency gambling. For example, ASRS #615534 reported a pilot who encountered an unusual attitude immediately after take off. The pilot was motivated to do the fight because he had a meeting to attend. The weather was reported as poor but the pilot thought that he could take off and then look at the weather to make his go/no go decision since he had observed this practice many times when flying with another pilot. He took off and went immediately into the clouds. The pilot panicked and became disoriented. He stalled the airplane, recovered, then entered a steep spiral at a low altitude.

The strategies above may help make the pilot more knowledgeable about the weather. Greater knowledge will lead to greater confidence and better responses when confronted with a weather situation.

CONCLUSION

The complexity of aeronautical decisions cannot be overstated. To train enhanced aeronautical decision-making (EADM), theory and knowledge from a variety of sources is imperative. Here we blend information from current research and theory in human learning, memory, decision making and expertise and apply it to database information, expert opinion, and a look at what is lacking in current training curriculum. These strategies are a part of an enhanced aeronautical decision making training tool being developed in the Beard laboratory. A prototype of this product will be available September 2005.

REFERENCES

[1] Ericsson, K.A., Krampe, R.T., and Heizmann, S. (1993) Can we create gifted people? In CIBA Foundation Symposium 178, *The origin and development of high ability*. (pp. 222-231). Chichester, UK; Wiley.

[2] Ericsson, K.A. (2001) The path to expert performance: Insights from the masters on how to improve performance by deliberate practice. In P. Thomas (Ed.) *Optimizing performance in golf*. (pp 1-57). Brisbane, Queensland, Australia: Australian Academic Press.

[3] Ettenson, R., Shanteau, J. and Krogstad, J. (1987) Expert judgment: Is more information better? Psychological Reports, 60, 27-238.

[4] Jacovone, J. and Dostal, M. (1992) A descriptive study of nursing judgment in assessment and management of cardiac pain. Advances in Nursing Science, 15, 54-63.

[5] Mosier, K.L. (1997) Myths of expert decision making and automated decision aids. In C. [2] Zsambok & G. Klein (Eds.) Naturalistic Decision Making. Hillsdale, NJ: Lawrence Erlbaum Associates.

[6] Schwartz, S. and Griffin, T. (1986) Medical thinking: The psychology of medical judgment and decision making. NY: Springer-Verlag.

[7] Shanteau, J. (1992) How much information does an expert use? Is it relevant? Acta Psychologica, 81, 75-86.

[8] Simon, H.A. and Chase, W.G. (1973) Skill in chess. *American Scientist, 61*, 394-403.

[9] Calderwood, R., Klein, G. A. and Crandall, B.W. (1988) Time pressure, skill and move quality in chess. *American Journal of Psychology*, *101*, 481-493.

[10] Simon, H.A. (1955) A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99-118.

[11] Orasanu, J., Martin, L. and Davidson, G. (2001) Cognitive and contextual factors in aviation accidents: decision making. In E. Salas & G. Klein (Eds.) Linking Expertise with Naturalistic Decision Making. Lawrence Erlbaum Ass. Inc.

[12] Standing, L. Conezio, J. and Haber, R.N. (1970) Perception and memory for pictures: Single trial learning of 2560 visual stimuli. *Psychonomic Science, 19*, 73-74.

[13] Fischer, U. and Orasanu, J. (2000) Do you see what I see? Effects of crew position on interpretation of flight problems. NASA Technical Memorandum #2000-[TK]. Moffett Field, CA: National Aeronautics and Space Agency.

[14] Klein, G. (1993) Sources of error in naturalistic decision making. In Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting (pp. 368-371). Santa Monica, CA: Human Factors and Ergonomics Society.

[15] Reason, J. (1990) Human Error. Cambridge, England: Cambridge University Press.

[16] Hollenbeck, J., Ilgen, D., Phillips, J., and Hedlund, J. (1994) Decision risk in dynamic two-stage contexts: Beyond the status quo. Journal of Applied Psychology, 79, 592-598.

[17] Ericsson, K. A. (2002) Attaining excellence through deliberate practice: Insights from the study of expert performance. In M. Ferrari (Ed.) *The pursuit of excellence in education*. (pp. 2-55). Mahwah, NJ: Lawrence Erlbaum Associates.

[18] Ericsson, K. A. (2005) Superior decision making as an integral quality of expert performance: Insights into the mediating mechanisms and their acquisition through deliberate practice. In *How professionals make decisions.* H. Montgomery, R. Lipshitz, B. Brehmer (Eds.) Mahwah, NJ, Lawrence Erlbaum Associates.

[19] Aviation Instructors Handbook

General Aviation Private Pilot Survey / Designated Pilot Examiner Program Assessment

Carla Hackworth, Ph.D.
Cristina Bates
Kali Holcomb, A. A.
Cristy Detwiler, B. A.
Civil Aerospace Medical Institute
Oklahoma City, OK

In order to meet objective 2 of the FAA Flight Plan, the FAA's General Aviation and Commercial Division (AFS-800) plans to improve the Designated Pilot Examiner (DPE) program. A DPE administers a practical test to evaluate a pilot's knowledge and skill to perform a flight maneuver. Problems arise when the application of DPE test criteria vary between examinations, examiners, or examinees. Two surveys will be conducted to determine the effectiveness of the examiner oversight program. One survey will be administered to all newly certified general aviation single-engine-land private pilots. In order to provide a balanced perspective of the practical test standard process of single-engine-land certification, we will also give DPEs the opportunity to comment on the process.

INTRODUCTION

In October 2004, the General Accounting Office released a report that was critical of the FAA's designee programs. FAA Flight Standards Service (AFS) is responsible for effective oversight of the Designated Pilot Examiner Program. Flight Standards District Offices (FSDO) provide direct oversight for this program. The FSDOs maintain DPE records, current Vital Information System data, and ensure compliance with other pertinent FAA policy orders. They must also ensure that DPEs conform to the pilot certification requirements of 14 CFR part 61 and the Practical Test Standards. DPEs conduct over 95% of all pilot practical tests.

Any appropriately qualified airman may apply to the FAA for designation. Any pilot examiner may, as authorized by his/her designation, accept applications for the practical tests necessary for issuing pilot certificates and ratings under 14 CFR Part 61. The examiner then conducts those tests and can issue pilot certificates and ratings to qualified applicants.

Ensuring that DPEs are providing complete and thorough practical tests to pilots is one of the many safeguards in place to maintain general aviation safety. That is, initially, flight instructors teach pilots the necessary knowledge and skills for safe flight operations. Following a demonstration of proficiency to their instructor, pilots must then demonstrate these skills to a designee of the FAA via a practical test. It is critical that DPEs are consistent in their test criteria across applicants. Additionally, DPEs must test all required elements of the practical test in compliance with FAA requirements. Unfortunately, this has not been found to be the case in all circumstances.

The FAA's Southwest Region found inconsistencies between FAA policy orders and part 61 requirements, on the one hand, and reported DPE practice on the other. Questionnaires sent to newly certified private pilots revealed that some applicants received incomplete practical tests. For example, respondents reported that examiners were repeating questions that were answered incorrectly by the pilots. Also, examiners were allowing the repetition of maneuvers that were performed poorly. Both of these actions do not conform to the Practical Test Standards.

Our purpose is to assess the extent to which the test criteria used by DPEs are in compliance with the pilot certification requirements of 14 CFR part 61 and the Practical Test Standards. Additionally, we will examine the consistency between examinations and examiners. To accomplish the research objectives, a general aviation private pilot national survey will be administered to all recently certified general aviation pilots across the United States. It is essential that newly certified general aviation private pilots should complete the survey as soon as possible after the practical test. Results of the survey will be used to determine the need for additional DPE training and/or oversight, identify areas of concern so that the FAA may affect corrections in FAA policy, guidance material, and FAA-sponsored programs in order to improve the overall quality of flight training and testing.

METHOD

Newly certified airplane single-engine land general aviation (GA) pilots across the United States will receive an anonymous and voluntary survey to complete and return by mail.

ASEL Survey

The survey includes over 40 questions that ask pilots about their flight training and practical test experience. There are several questions about the pilot's experience with their independent flight instructor or pilot school. Respondents are asked for example: "In preparation for your practical test, did your flight instructor advise you that the pilot examiner would assess runway incursion avoidance?" Respondents are also asked about their examiner and their most recent practical test. For example, respondents would indicate yes or no to, "Did the examiner ask you to explain the maintenance logbook entries for the aircraft you used during the practical test?" Additionally, respondents indicate yes or not to which of the technical subject areas, maneuvers, and procedures the examiner required them to explain, demonstrate, or repeat. Although not exhaustive, some areas included: Preflight Procedures, Takeoffs, Landings, and Go-Arounds, Stalls, Maneuvers/Procedures in Simulated Instrument Conditions, and Emergency Operations. Respondents are also asked about specific critical maneuvers. For example, "On your most recent practical test, did you demonstrate a crosswind landing?"

DPE Survey

A Designated Pilot Examiner survey will be administered to all DPEs, of which there are approximately a thousand, across the United States. Names, addresses, and respective FAA regions of each DPE were obtained from AFS-900.

The survey includes items that ask DPEs about their practical testing procedures and practices. DPEs are asked how many Private Pilot Airplane Single-Engine-Land (ASEL) practical tests they conducted during the past 12 months, the percentage of first-time Private Pilot ASEL applicants that perform unsatisfactorily during the flight portion of the practical test, and if they use a written plan of action (POA) when conducting a practical test for the Private Pilot ASEL rating.

Additionally, DPEs are asked about the current level of proficiency amongst their applicants. Specifically, they are asked how adequately instructors are preparing first-time Private Pilot ASEL applicants for the oral (ground) portion of the practical test.

The Civil Aerospace Medical Institute (CAMI) Aerospace Human Factors Division will mail all ASEL DPEs an anonymous and voluntary survey with a postage-paid return envelope. In an attempt to maximize response rates, the survey will have an attached cover letter explaining the purpose of the survey and asking for feedback regarding DPE examination practices. Respondents will be assured that the survey is completely anonymous and voluntary and that if any of the questions make them feel uncomfortable, they should skip them.

We wish to capture the opinions and attitudes of respondents soon after their practical test flight. Thus, it will be necessary to conduct multiple data collections because GA pilot certification testing occurs repeatedly throughout the year.

Returned surveys will be scanned into a database through the use of Teleform software. Summary reports for the surveys will be created for each region that has at least 8 respondents. In addition, an overall report will be created.

RESULTS

Three thousand GA ASEL surveys and envelopes were printed and are ready for distribution. AFS-800 has asked the registry to assist with the ASEL survey project. The registry is to gather the pertinent names and addresses for the ASEL GA pilots and supply the addresses to CAMI personnel. However, there a few criteria that they cannot match with our original target population (e.g., region of respondent identified, first-time applicant with no failures in the past); therefore, the project is on hold waiting for direction from AFS-800.

REFERENCES

General Accounting Office (2004). FAA needs to strengthen the management of its designee programs. GAO-05-40.

HOW HIGH IS HIGH ENOUGH? QUANTIFYING THE IMPACT OF AIR TRAFFIC CONTROL TOWER OBSERVATION HEIGHT ON DISTANCE PERCEPTION

[1]William K. Krebs, Glen Hewitt
Federal Aviation Administration, Washington, D.C.
Steven R. Murrill
Army Research Laboratory, Adelphi, Maryland
Ronald G. Driggers
US Army RDECOM CERDEC NVESD, Ft. Belvoir, Virginia

Each year the Federal Aviation Administration (FAA) builds approximately seven air traffic control towers in the national airspace system. Each airport has unique surface and airspace characteristics, but all airports must determine the location and height of the new air traffic control tower (ATCT). These two factors impact cost and safety, therefore the FAA must develop a quantitative means in measuring what improvement in ATCT visibility can be gained by increasing tower height at different locations on the airport surface. Two metrics were developed (Object Discrimination, Line of Sight Angle of Incidence) to assess the impact of tower height on distance perception. The two metrics are fairly robust and easy to use to assess the impact of tower height on air traffic control tower specialist distance perception.

Introduction

"The air traffic control tower siting process must take into consideration criteria relating to the safety of air traffic operations for each site. The optimum height and location is the result of balancing many requirements and considerations, based on the current approved Airport Layout Plan (ALP). The goal of this process is to maximize operational performance and safety when siting an ATCT. (6480.xx, page 3)".

A Federal Aviation Administration employee requested assistance in determining a proposed tower height. The employee's request stated:

> *"I've been asked to justify a certain height at a new tower. I've tried to explain to the Terminal Business folks that this place needs a taller tower because of line of sight problems, heat wave distortion, night time glare from lighting that surrounds the airport, and a parallax type of problem when watching aircraft approaching the airport for landing on closely spaced parallel runways. (FAA employee, 2004)"*

The Federal Aviation Administration William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory tower cab simulation allows air traffic control tower specialists to assess the impact of a proposed tower height and location on an airport surface. The AFTIL can simulate real-world scenes to assess the physical attributes of the tower cab relative to the airport surface and how that may affect visibility, such attributed include cab orientation, tower look-down angle, look across line-of-site, mullions, look-up angle for missed approaches, movement and non-movement areas; unobstructed views. The diversity of the AFTIL has tradeoffs; specifically to depict a real-world scene in a 360^0 tower cab simulation spatial resolution of the generated scene is sacrificed due to amount of computer processing required to generate a scene. In normal mode, the AFTIL image generated scene is equivalent to 20/80 visual acuity which is more than sufficient to address the most of the tower siting criteria. However, the AFTIL can not address the impact of tower height on an air traffic control tower specialists' detection of a distant object.

The objective of this study was to develop, test, and validate a set of human performance metrics to assess the impact of tower height on air traffic control tower specialist distance perception. The human factors metrics as well as the AFTIL simulation will be used to site a tower at an airport.

Methods

Object Discrimination

Question: What improvement in detecting or recognizing a distant object can be gained by increasing tower height or decreasing tower distance from the object?

[1] To whom correspondence should be addressed. Current address: Federal Aviation Administration, 800 Independence Avenue, ATO-P R&D, Room 907A, Washington D.C. 20591 or william.krebs@faa.gov, 202-267-8758.

The overall objective is to provide the FAA with a user-friendly software tool that provides quantitative information on the impact of ATCT height on aircraft visibility. The tool includes drop-down windows for user input as well as graphical chart windows for results output. The primary output of this tool is probability-of-discrimination (detection and recognition) curves as a function of observation range and tower height. The tool draws from four well-developed and empirically-validated functions and models: The U.S. Army Night Vision Laboratory's Standard Target Transfer Probability Function (using modified Johnson's discrimination criteria), Barton's model for the human eye's Contrast Transfer Function, Kopeika's atmospheric (optical) turbulence modulation transfer function, and Tatarski's atmospheric-index-structure-parameter height-scaling model. In addition, the algorithms and routines include two enhanced-accuracy features that account for: the impact of turbulence on a downward-slanting optical path, and the effect of distance between the point of optical path integration and the observer (the "shower curtain" effect).

Model Assumptions:

(a) Detection is defined as the ability to notice the presence of an object on the airport surface without regard to the class, type, or model (e.g., an object such as an aircraft or vehicle). The observer knows something is present but cannot recognize or identify the object.

(b) Recognition is defined as the ability to discriminate a class of objects (e.g., a class of aircraft such as single engine general aviation aircraft).

(c) The object (aircraft or vehicle) size is taken to be the square root of the frontal or side cross-sectional area of the object (e.g., wing span x height).

(d) Modified Johnson's criteria is used for the number of optical cycles required for a 50% probability of success in object discrimination (N50).

(e) All observations are made with the unaided eye.

(f) The observer is assumed to be at the specified tower height while all objects (e.g., aircraft, vehicles) are taken to be at the ~ 3 ft (1 m) height.

To account for the impact of atmospheric (optical) turbulence on the downward-slanting optical path, an average/effective refractive-index-structure-parameter *scaling factor* was calculated. This *scaling factor* was derived by taking the line integral of the Tatarski height scaling equation over the downward-slanting optical path.

Object Discrimination Tool: The tool (figure 1) can be found at http://www.hf.faa.gov/visibility.

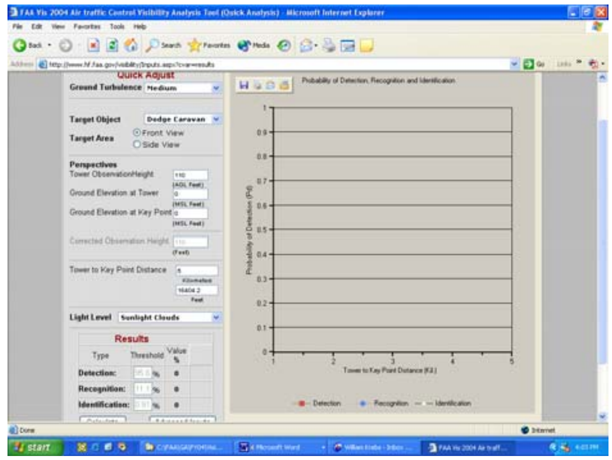


Figure 1. Object discrimination tool graphical user interface. Users enter tower height and distance to calculate air traffic control tower specialists detection and recognition of an airport surface object.

Procedure: From the graphical user interface select object, specify tower height and key point distance, specify ground turbulence, and specify outside illumination level. Key point distance is defined as the distance between the air traffic control tower and object of interest on the on the airport surface.

Results: Probability of detection and recognition values were calculated for one hundred and ninety five level seven or greater air traffic control towers in the national airspace. Key point was defined as the most distant runway threshold from the air traffic control tower for each airport. The object was a front-view of a Dodge Caravan minivan set at 33% contrast. Illumination was sunlight clouds and ground turbulence was dependent upon geographical location.

Based on the 195 air traffic control tower sample, criterion was set at 1½ standard deviations below the sample mean (i.e., better than 6.7% of the sample) which is equivalent to 95.5% for detection and 11.5% for recognition (table 1).

Observation Capability Requirements	Observation Description	Front View Probability Criteria Minimum
Detection	Ability to notice the presence of an object on the airport surface without regard to the class, type, or model (e.g., an object such as an aircraft or vehicle). The observer knows something is present but cannot recognize or identify the object.	95.5%
Recognition	Ability to discriminate a class of objects (e.g., a class of aircraft such as single engine general aviation aircraft).	11.5%

Table 1. Probability of discrimination detection and recognition criterion values based on one hundred and ninety five level seven or greater air traffic control towers in the national airspace.

Line of Sight Angle of Incidence
Question: What improvement in the controller's viewing perspective can be gained by increasing the observer's line of sight angle of incidence to the airport surface at key distance points?

Observers: Twelve tower-rated air traffic control specialists, age 26-59 years, were recruited from four different tower airport facilities. Average air traffic control tower experience was 17.4 years. All observers had normal or corrected-to-normal visual acuity, and had normal color vision. All observers granted informed consent prior to participation. All observers were naïve to the experimental hypothesis.

Apparatus: Federal Aviation Administration William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory's (AFTIL) nine Quantum 3D "Alchemy" image generators (IGs) drove nine, six-foot vertical by eight-foot horizontal rear-projection screens arranged in a 360^0 circular pattern to simulate an air traffic control tower cab environment. The diameter of the simulation floor plan is 24'. Each rear-projector, Epson "PowerLight" model 9100, had a pixel resolution set at 1280 (horizontal) by 1024 (vertical) pixels with a field-of-view of approximately 20^0 (horizontal) by 15^0 (vertical). To increase resolution of the visual simulation, three of the nine rear-projection screens were used in the test. Observers were positioned 24' from the most distant screen thereby allowing a resolution of 64 pixels per degree. The base of the screens is approximately 30 inches from the floor to allow an average standing observer's eye-height to be centered on the screen. Software used to model the simulation were AutoCad, MultiGen-Paradigm, PhotoShop, and other graphic simulator tools to generate vehicle ground and air routes for the airport. Frame rate was fixed at 30 frames/second.

Airport Display: The AFTIL tower simulation displayed a realistic depiction of an airport surface using panoramic photographs and computer graphics (figure 2). The visual simulation contained terrain features, hangers, terminals, runways, taxiways, as well as dynamic surface and airborne aircraft and other ground surface vehicles.

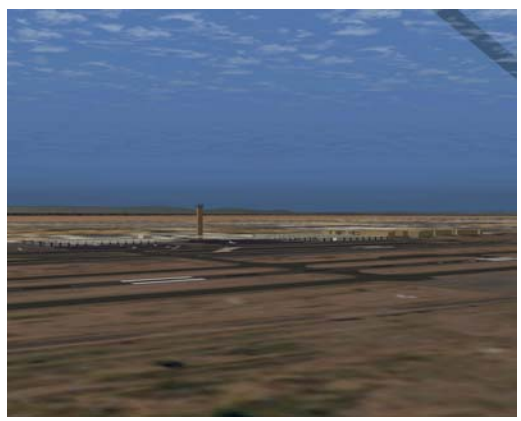


Figure 2. Simulated air traffic control tower scene generated by the Federal Aviation Administration William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory.

Eight ATCT simulations were created: Cahokia/Saint Louis Downtown (CPS), Fort Wayne International (FWA), New York/La Guardia (LGA), Memphis International (MEM), Morriston Muni (MMU), Minneapolis-Saint Paul International (MSP), Oshkosh/Wittman Regional (OSH), and Richmond

International (RIC). At each airport, a critical key point was selected. Observers were informed on the location of the key point. All simulations were displayed during day illumination.

<u>Procedure:</u> The observer was exposed to fifty experimental dynamic scenes: five of eight ATCT simulations and ten tower observation heights (table 1). In each trial, observers performed common air traffic control tower visual tasks at different tower heights. The observer's task was to visually scan a designated distant "key point" on an airport surface and rate the ability to (1) distinguish boundaries of the movement areas and (2) identify position of target at the airport's key point. The distant "key point" was an MD-80 located on the airport surface. Prior to entering the tower cab simulation, the experimenter familiarized the observer to a 6-point Likert rating scale and the response criteria for each question. At the beginning of each block of trials, observers were afforded several minutes to familiarize themselves with the airport layout and location of the distant key point. At the completion of the familiarization, the observer's eyes were occluded and the first experimental tower height was selected. The experimenter then instructed the observer to open his or her eyes and respond to both questions. Within each block of trials, tower height was randomly assigned without replacement. At the completion of the tenth tower height, the next ATCT scene was presented and the same procedure was repeated. ATCT scene order was randomly assigned across observers. Reaction time was not recorded.

<u>Results:</u> Calculate the height of the observer in the tower according to the formula:

$$H_O = (H_C - (P_E - T_E)),$$

where, H_O is height of observer; H_C is controller eye height; P_E is ground elevation of key point Above Mean Sea Level; T_E is ground elevation of tower Above Mean Sea Level. Controller eye height is defined as five feet above cab floor height.

Compute the Line of Sight angle at which the observer's view intersects with the airport surface at the key point.

> Line of Sight angle = ArcTan (height of observer/distance between key point and tower)

Based on the responses of twelve observers and several other air traffic tower controller specialists, the minimum level of performance for question 1 (*How well can you distinguish boundaries of the movement areas?*) was response 2 (*Can discriminate boundaries of most of runways and taxiways; but provides no distance information*). Figure 3 illustrates observers' proportion of "yes" responses for response of 2 or greater. All observers reported a response of 2 or greater when towers line of sight angle of incidence was 1.5 degrees or greater. Converting proportion of "yes" responses for response 2 or greater to Z scores then fitting a linear line showed that 50% of the observers reported 0.481 degrees as the preferred line of sight angle of incidence (figure 4).

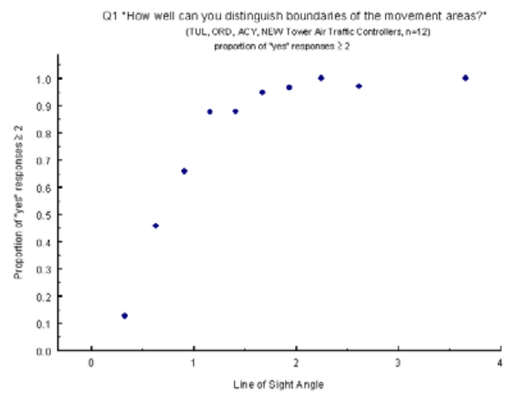


Figure 3. Illustrates observers' proportion of "yes" responses for response of 2 or greater for question "*How well can you distinguish boundaries of the movement areas?*"

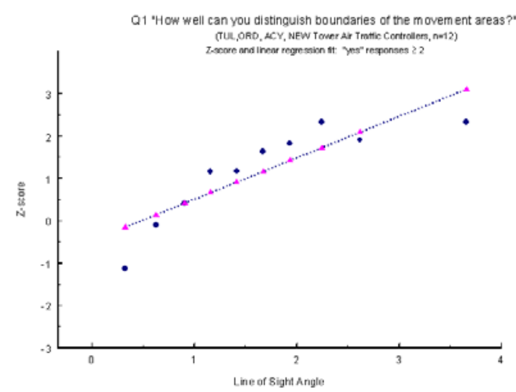


Figure 4. Converting proportion of "yes" responses for response 2 or greater to Z scores then fitting a linear line showed that 50% of the observers reported 0.481 degrees as the preferred line of sight angle of incidence.

For question 2 (*How well can you identify the position of an object relative to the airport's key point?*), the minimum acceptable response was 3 (*Able to determine that object position is in general vicinity of key point, but unable to estimate distances of object within movement area*). Figure 5 and 6 illustrate observers' responses for a response of 3 or greater and linear fit to Z scores, respectively. Fifty percent of the observers reported 0.799 degrees as the preferred line of sight angle of incidence (figure 6).

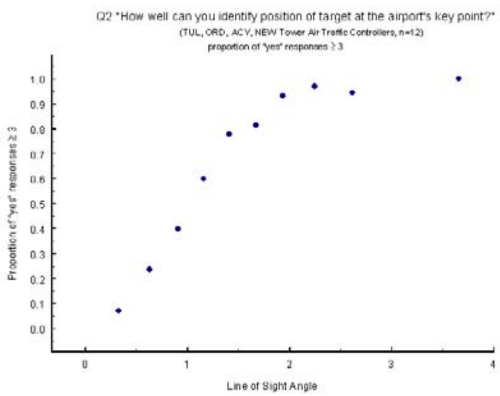


Figure 5. Illustrates observers' proportion of "yes" responses for response of 3 or greater for question "*How well can you identify the position of an object relative to the airport's key point?*"

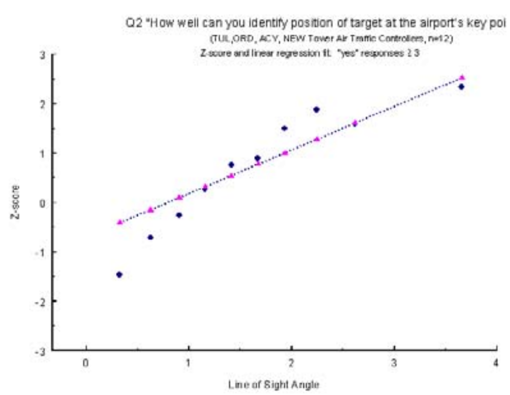


Figure 6. Observers reported 0.799 degrees as the preferred line of sight angle of incidence for a response of 3 or greater.

The minimum line of sight angle of incidence is set at 0.799. The higher value was selected due to question 2 was reported as the more important task of an air traffic control tower specialist.

Conclusions

The analyses performed may assist air traffic requirements in determining future air traffic control tower heights. To assist the decision team, the analyses could be plotted to illustrate percent improvement of air traffic control tower specialists' recognition or identification of an aircraft by tower height expressed in dollars per linear foot. Of course, there are many factors that determine tower height and location but the analyses described above may provide air traffic requirements additional quantitative data to assist in their decision.

References

Federal Aviation Administration (November 10, 1972). Department of Transportation, Federal Aviation Administration Publication Order #6480.4, Airport Traffic Control Tower Siting Criteria.

Federal Aviation Administration (in preparation). Department of Transportation, Federal Aviation Administration Publication Order #6480.xx, Airport Traffic Control Tower Siting Process.

Acknowledgements

Sponsored by Federal Aviation Administration, Human Factors Research and Engineering Division. Also, sponsored by Bill Kaplan, Federal Aviation Administration, Terminal Facilities.

Thanks to Scott Iwamoto, Kathy Peterson, Dave Cloutier for their never ending support in allowing us to participate in the air traffic control tower siting process, and a special thanks to Bill Vaughan for providing access to the William J. Hughes Technical Center Airway Facilities Tower Integration Laboratory. We also thank John Aschenbach, Rodman Bourne, Roger Bawgus, and Bernie Garbowski in their assistance with the simulation data collection, and Rob Heckart and Doug Norvell for placing the tower visibility tool on the website.

The views expressed in the article are those of the authors, and do not reflect the official position of the Department of Transportation, Department of the Navy, the Department of Defense, or the United States Government.

Human Error Associated with Air Medical Transport Accidents in the United States

Albert Boquet
Embry Riddel Aeronautical University
Daytona Beach, FL

Scott Shappell
Clemson University
Clemson, SC

Kali Holcomb, Cristy Detwiler, Cristina Bates, Carla Hackworth
Civil Aerospace Medical University
Federal Aviation Administration

Doug Wiegmann
Mayo Clinic
Rochester, MN

Helicopter emergency medical services (HEMS) play a vital and growing role in the U.S. healthcare industry. However, since 1998, there has been a troubling increase in the number of accidents associated with this group. Similar to data for other aircraft, the majority of these accidents are human error related. This investigation used the Human Factors Analysis and Classification System to categorize human error in HEMS operations. Like other aviation operations, skill-based errors comprised the majority of the unsafe acts, followed by decision errors, violations and perceptual errors. Also troubling was the number of fatalities associated with weather and night-related accidents, as well as controlled flight into terrain.

On January 11, 1998, near Sandy, Utah, an air ambulance was attempting to evacuate an injured skier when it impacted a ridge shortly after take off killing all on board. Witnesses reported blizzard conditions with wind gusts up to 35 knots, and significantly reduced visibility. Causes cited for the crash are an all too familiar litany of human error, including flight into known adverse weather, failure to maintain clearance, darkness, heavy snow, high winds, and perceived pressure to fly.

A 1966 report by the National Research Council of the National Academy of Sciences identified accidents as the leading cause of death for persons between the ages of 1 and 37, while listing it as the fourth leading cause of death for all ages. Based upon battlefield statistics which revealed a direct correlation between survival of battlefield wounds and prompt evacuation and treatment, the council recommended the use of helicopters for the transport of critically ill or injured patients to trauma centers that were equipped to handle these cases. This effort was further supported by the concept of the "The Golden Hour" which refers to the reduction in morbidity and mortality that results from immediate treatment of trauma victims (Cowley, 1976).

The first privately funded hospital-based helicopter program was established at St. Anthony's Hospital in Denver, CO in 1972 (Thomas, 1988). Since that time, helicopter emergency medical systems (HEMS) operators have undergone tremendous growth. In fact in 2001, according to the Association of Air Medical Services there were more than 300,000 patients transported.

Unfortunately, this growth has not come without some problems. A spate of accidents during a period of rapid growth from the early to mid 1980's raised some initial red flags. From 1980 to 1987 there were 54 accidents or an average of 7.7 accidents per year. This rate improved from 1988 to 1997 revealing an average accident rate of 4.9 accidents per year. Similar to other reports, (Blumen, 2002) we found that since 1998 there has been a steady and alarming increase in the accident rate of HEMS (Figure 1). Like other areas of aviation, the primary cause, cited in over 70% of the accidents, was human error. In fact, Blumen goes on to state that the most common factors included weather and dark night conditions.

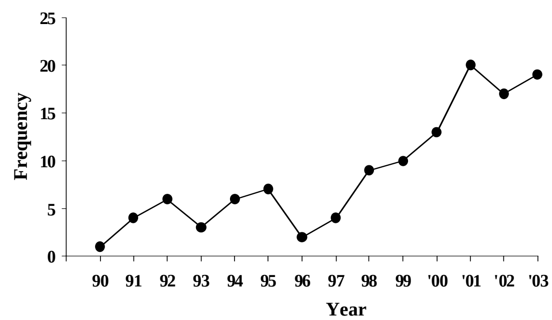


Figure 1: HEMS accidents by year

While discussions of accident rates, human error percentages, and environmental conditions may shed some light on the situation, they do not, in and of themselves provide the clarity necessary, to fully address the rise in accident rates. For that reason, an analysis of all HEMS accidents from 1990 to 2003 involving human error was conducted using the Human Factors Analysis and Classification System (HFACS).

HFACS

The entire HFACS framework includes a total of 19 causal categories within Reason's (1990) four levels of human failure. While in many ways, all of the causal categories are equally important; particularly germane to any examination of HEMS accident data are the unsafe acts of aircrew. For that reason, we have elected to restrict this analysis to only those causal categories associated with the unsafe acts of HEMS aircrew. A complete description of the HFACS causal categories is therefore beyond the scope of this report and can be found elsewhere (Wiegmann & Shappell, 2003).

Unsafe Acts of Operators

In general, the unsafe acts of operators (in the case of aviation, the aircrew) can be loosely classified as either errors or violations (Reason, 1990). Errors represent the mental or physical activities of individuals that fail to achieve their intended outcome. Not surprisingly, given the fact that human beings by their very nature make errors, these unsafe acts dominate most accident databases. Violations, on the other hand, are much less common and refer to the willful disregard for the rules and regulations that govern the safety of flight.

Within HFACS, the category of errors was expanded to include three basic error types (skill-based, decision, and perceptual errors). In general, decision errors represent conscious decisions/choices made by an individual that are carried out as intended, but prove inadequate for the situation at hand. In contrast, skill-based behavior within the context of aviation is best described as “stick-and-rudder” or other basic flight skills that occur without significant conscious thought. As a result, these skill-based actions are particularly vulnerable to failures of attention and/or memory as well as simple technique failures. Finally, perceptual errors occur when sensory input is degraded or “unusual,” as is often the case when flying at night, in the weather, or in other visually impoverished conditions.

As with errors, there are many ways to distinguish between types of violations. However, two distinct forms are commonly referred to, based upon their etiology. The first, routine violations, tend to be habitual by nature and are often tolerated by the governing authority. The second type, exceptional violations, appear as isolated departures from authority not necessarily characteristic of an individual’s behavior nor condoned by management.

METHOD

The National Aviation Safety Data Analysis Center (NASDAC) and NTSB were utilized to identify human-error-related HEMS accidents, specifically medical flights operating under 14 Part 91 (ferrying or repositioning flights) and 14 Part 135 (patient transport). This resulted in 121 accidents, as reported by the National Transportation Safety Board (NTSB) from 1990 to 2003. For the purpose of this report, we decided to limit the investigation to only those accidents (N=74) occurring in what we have termed the “rescue triangle” (see Figure 2). In addition, training accidents, fixed wing, and maintenance repositioning flights were also eliminated.

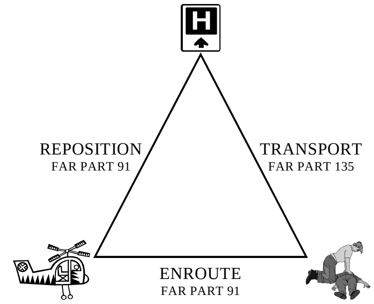


Figure 2: EMS Rescue Triangle

Subject Matter Experts

Six GA pilots were recruited from the Oklahoma City area as subject matter experts and received roughly 16 hours of training on the HFACS framework. All six were certified flight instructors with a minimum of 1,000 flight hours in GA aircraft (mean = 3,530 flight hours) as of June 1999. After training, the six GA pilot-raters were randomly assigned accidents so that two separate pilot-raters analyzed each accident independently. Using narrative and tabular data obtained from both the NTSB and the FAA NASDAC, the pilot-raters were instructed to classify each human causal factor using the HFACS framework. Note, however, that only those causal factors identified by the NTSB were classified. That is, the pilot-raters were instructed <u>not</u> to introduce additional casual factors that were not identified by the original investigation. To do so would be presumptuous and only infuse additional opinion, conjecture, and guesswork into the analysis process.

After our pilot-raters made their initial classifications of the human causal factors (i.e., skill-based error, decision-error, etc.), the two independent ratings were compared. Where disagreements existed, the corresponding pilot-raters were called into the laboratory to reconcile their differences and the consensus classification was included in the database for further analysis. Overall, pilot-raters agreed on the classification of causal factors within the HFACS framework more than 85% of the time, an excellent level of agreement considering that this was, in effect, a decision-making task.

RESULTS

As Figure 3 illustrates, human error accounts for the lions’ share of the accidents in the HEMS population, as it does in all categories of aviation. Figure 3 illustrates the overall number of HEMS accidents, compared to those human error related accidents occurring in the rescue triangle only.

If one examines the characteristics of the curve, there is a steady increase in the number of accidents beginning in 1998, which is maintained through 2003.

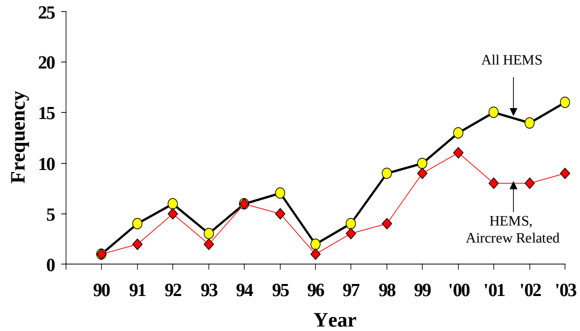


Figure 3: Human Error Associated HEMS Accidents

Unsafe Acts of the Operators

After applying HFACS codes to these data, the types of human error involved displays a familiar pattern (Figure 4), characterized by more skill-based errors (59.5%), followed by decision errors (33.8%), then perceptual errors (18.9%), and violations (14.9%). However, this in and of itself does not provide the resolution necessary to adequately assess what these unsafe acts means for aircrew operations.

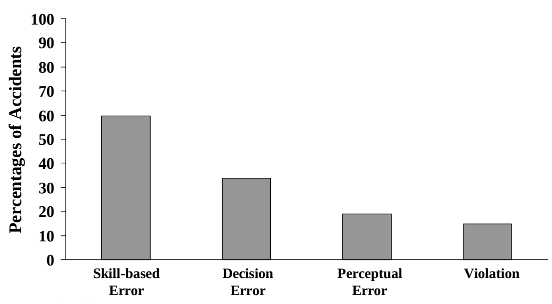


Figure 4: Unsafe Acts of the Aircrew

Fine-Grained Analysis

In order to better understand the types of errors made by HEMS aircrews, a fine-grained analysis was conducted to identify the specific types of unsafe acts that were associated with the HEMS accidents. The top three errors for each unsafe act are reported here. For skill-based errors, the most common errors were, failure to maintain clearance (28.6%), aircraft control, visual lookout, and altitude/clearance (8.2%). The top decision errors were in-flight planning/decision making and unintentional VFR flight into IMC (both 17.9%), followed by remedial action (10.7%). Perceptual errors consisted of aircraft control (25%), followed by distance/altitude and altitude/clearance (both 12.5%). Finally, the top violations were all weather related, including procedures and directives not followed (30.8%), VFR flight into IMC and flight into known adverse weather (15.4%).

As these data indicate, for skill-based errors, clearance from objects and terrain make up the bulk of the errors. On the other hand, decision errors, perceptual errors, and violations overwhelmingly were made up of errors that occurred in degraded conditions, either weather or night operations.

Relationship of Unsafe Acts to Fatalities. When the relationship of unsafe acts to fatalities was examined, the data revealed that those accidents associated with violations claimed a higher percentage of lives, compared to the other categories of unsafe acts. Of those accidents involving a skill-based error, 31.8% resulted in a fatality, compared with 20% for decision errors, and 42.9% for perceptual errors. However, when a violation was involved, 63.6% of these accidents had at least one fatality (Figure 5). This is consistent not only with other flight deck operations, but with maintenance violations as well.

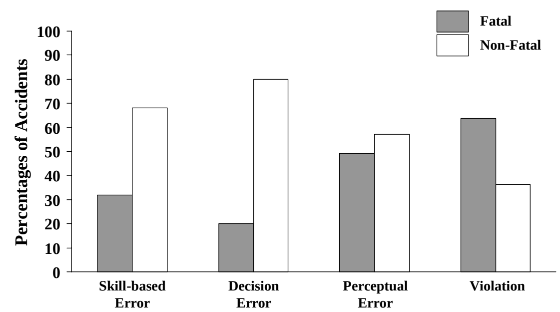


Figure 5: Fatalities Related to Unsafe Acts

Accidents by Position. When the data were analyzed by position alone (enroute, transport, or reposition), the largest number of accidents was found in the enroute phase, followed by reposition, then transport. This was expected since one may assume more pressure to arrive at the pickup site as quickly as possible (Figure 6).

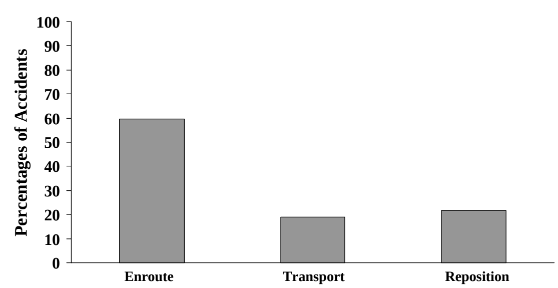


Figure 6: Percentage of Accidents by Position

Phase of Flight. Because of the unique nature of the HEMS flight operations, an analysis of phase of flight by position was conducted. During the enroute portion of the rescue triangle, the greatest number of accidents occurred during cruise and landing (both 20.5%), followed by approach and maneuvering (both 17.9%). Takeoff accounted for 12.8% followed by hover, taxi, and emergency descent (5.1%, 2.6%, and 2.6% respectively). For the transport phase of the triangle, the bulk of the accidents occurred during the takeoff phase of flight with 58.3% of the accidents occurring while leaving the scene. The remaining accidents were evenly distributed during climb, descent, approach, go-around, and maneuvering, all with 8.3% of the accidents. Finally, during reposition, the majority of the accidents occurred during cruise flight (46.2%), followed by takeoff (15.4%). The remaining accidents were divided by standing, descent, approach, landing, and emergency landing after takeoff, with 7.7% of the accidents occurring in each of these phases.

Lighting Conditions. In order to determine the effects of time of day on HEMS operations, the data was divided into day vs. night operations. Furthermore, rather than simply compare accident rates associated with time of day; fatalities associated with time of day were also analyzed (Figure 7).

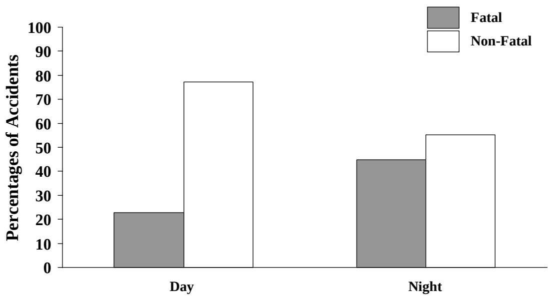


Figure 7: Fatalities by time of day

While the actual percentages of accidents associated with time of day were relatively evenly split between day and night, 47.9% and 52.1% respectively, the breakdown in fatalities was not. As can be noted in Figure 7, 22.9% of daytime accidents were associated with fatalities compared with 44.7% fatalities when the accident occurred at night.

Weather Conditions. The vast majority of HEMS accidents occurred in VMC weather (74.3%) vs. IMC (25.7%) However, similar to the analysis of lighting conditions, there were more likely to be fatalities associated with IMC weather (Figure 8). When examining the relationship between fatalities and weather conditions, IMC operations took a greater toll with 73.7% resulting in fatalities, compared with only 20.0% of VMC related accidents resulting in fatalities. To better illustrate this point, the odds of dying in an accident in IMC weather are 11 times greater when compared to VMC conditions.

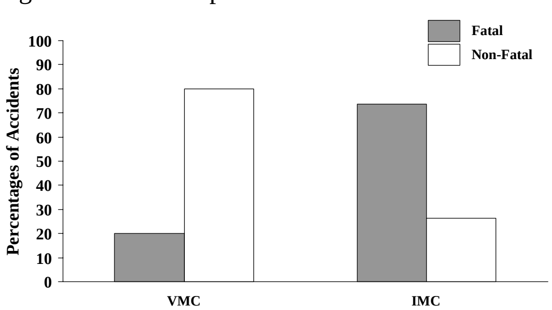


Figure 8: Fatalities related to weather conditions

Controlled Flight into Terrain/Obstacle. Because there were 25-controlled flights into terrain (CFIT) in this population, a closer look was called for. For this analysis, CFITs were broken down into two categories, controlled flight into terrain (CFIT/T) and controlled flight into obstacle (CFIT/OBS). In order to gauge the effects of degraded conditions on the occurrence of CFITs, an impoverished variable was created by combining both night/dusk conditions or poor weather creating an impoverished variable. The results are displayed in Figure 9. As can be seen by the graph, the likelihood of CFIT/T greatly increases in impoverished conditions.

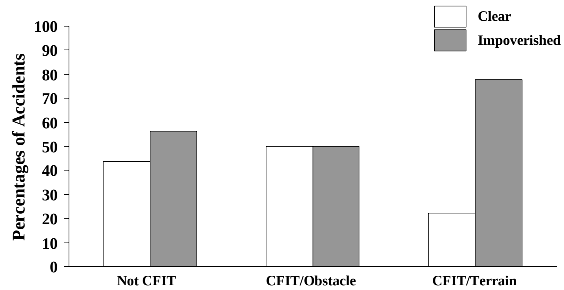


Figure 9: CFIT to Environmental Conditions

A similar analysis was carried out to determine the relationship of the different types of CFITs to fatalities. As Figure 10 illustrates, there was an increased likelihood of a fatality in a CFIT/T compared to CFIT/OBS (88.9% vs. 31.3%).

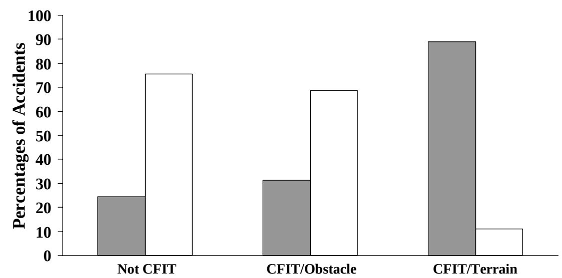


Figure 10: Relationship of CFIT to Fatal (Solid bars) vs. Non-Fatal (Clear bars) Accidents.

DISCUSSION

Human error involved in HEMS accidents was classified using HFACS, allowing for a higher level of description than is typically associated with standard reporting. For the purpose of this investigation, only those human error-related accidents occurring within what has been termed the rescue triangle have been analyzed. This was done in an attempt to capture what is a true HEMS operation. While it is understood that these do not make up all of the theaters within which HEMS flights operate (training, maintenance ferry flights, etc…) it was believed that the unique pressures involved in these operations are reflected within these limits.

In reviewing the data for unsafe acts, we see patterns similar to other aviation platforms. Skill-based errors were the most common type of human error in HEMS accidents, followed by decision errors, then perceptual errors and violations. Notably, however is the greater number of fatalities associated with violations when compared to the other unsafe acts. Those accidents, which involved violations, were 3 times more likely to be associated with a fatality. Similar data are reported in other studies using the HFACS (Shappell and Wiegmann, 2003).

Whereas the unsafe acts resemble other aviation platforms, the fine-grained analyses revealed some important differences. For instance, the observation that the top three violations were weather related speaks volumes for what are VFR operations, and what has plagued HEMS operations over the years. This is consistent with other reports, which cite weather as a problem in HEMS operations (Frazer, 1999). While it may be easy to blame the pilot who ultimately has the go/no go decision, one must also keep in mind that many have to rely on local weather forecasts, which may lack the detail necessary for pilots to make informed decisions in questionable weather. Thus, pilots may takeoff expecting the back door to be open for a return, only to find it quickly closed. Furthermore, pressure to fly, either self-induced because of the nature of the operation or induced by pressure to generate revenue, must be factored into the decision-making process.

When the position of the flight within the rescue triangle was considered, the greatest number of accidents was shown to occur in the enroute phase. This is not surprising, since this is the time when one may assume (although it is only an assumption) that the greatest pressure to "get there" may be present. While transport and reposition were not significantly different, it was somewhat surprising that slightly more accidents were occurring during the reposition phase. Frazer (1999), states that "get home itis" may be involved in these types of accidents. However, it may also be that during this phase, the aircrew may experience some complacency since the emergency is passed, thus, they may be somewhat less vigilant in the cockpit.

Phase of flight revealed that during the enroute phase of operation, cruise and landing were most problematic. Landing is best explained by the fact during this phase, the aircraft are landing at unimproved sites, often during less than ideal conditions. Cruise is somewhat more difficult to explain except that here is where pressure to arrive at the scene in the least amount of time may be at work. For transport, takeoff was by far associated with the most accidents. Again, the fact that aircraft are taking off from unimproved sites, contending with wires, fences, trees, etc… may best explain these data. Finally, during reposition, cruise accounted for almost 50% of the accidents. It is here that complacency and "get home itis" may factor into these accidents.

Consistent with the findings from Frazer in 1999, there was no difference in the number of accidents occurring during the day when compared to night operations. However, we took the analysis one step further, and found that those accidents occurring at night were almost twice as likely to be associated with a fatality. This should certainly be cause for scrutiny of the nighttime VFR operation that HEMS flies under.

The relationship for VFR vs. IMC conditions is even more lopsided. While the vast majority of accidents occurred in VMC conditions, the fatality rate associated with IMC related accidents was almost three times greater. Thus, flying in degraded conditions, whether due to darkness or poor weather, the chances of an accident do not necessarily go up, however, should an accident occur, the probability of a fatality greatly increases. Based upon these findings, accidents in poor weather, as well as darkness are costly indeed.

Nowhere is this more evident than in CFIT accidents. In comparing CFIT/OBS to CFIT/T, there were 16 CFIT/OBS with 8 occurring in clear conditions, and 8 occurring in impoverished. For CFIT/T, there were 9 accidents with 2 occurring in clear conditions and 7 occurring in impoverished conditions. Thus, for CFIT/T, there was over three times the number of accidents in degraded conditions. To make matters worse, of those 9 CFIT/T accidents, 8 included a fatality, compared with 5 for CFIT/OBS. This computes to a 2.5 times greater risk of a fatality if an aircrew is involved in a CFIT/T vs. CFIT/OBS.

So where does this leave the HEMS? While it is easy to sit back and in retrospect "arm chair quarterback" an industry that has become a mainstay of emergency medicine, the answers will not be as simple as they seem. Number one on the list to be addressed are operations in degraded conditions. The obvious recommendation here are IMC equipped aircraft and pilots who are truly instrument certified. While this may seem counterintuitive for an industry that operates under VFR rules, the number and severity of accidents that occur in weather and in night conditions coupled with the number of weather-related violations indicates that the time has come to consider IFR currency and similarly equipped aircraft.

Another solution that has been batted about are dual crew and dual engine aircraft. However, this presents problems for smaller operations, due to the increased expense of these aircraft and higher costs associated with additional crewmembers. Night vision goggles (NVGs) have also been suggested due to the number of accidents that occur at night. However, this is not supported by the data. Specifically, there were no more accidents at night compared to daytime operations. And while the severity of accidents occurring at night are greater in terms of fatalities, most of the nighttime accidents occurred in IMC, where NVGs would have been of no use Furthermore, NVGs do not increase visibility of wires and fence lines which pose problems at landing and take-off sites.

Finally, training for on-scene responders should be standard operating procedure. Law enforcement and ground crews should be educated as to where helicopters can safely land and take off from. They should be aware of what HEMS crews can and cannot see from the air, how much room they need to land and maneuver, how soft the ground can be before there is a problem, etc… This may help to decrease many of the landing and takeoff accidents noted in the enroute and transport phase of the operation.

It should be understood that any operation has a certain amount of risk associated with it, and HEMS operations are no different. But there are two ways to go about reducing this risk. One is to make the operation safer, in other words, to reduce the probability of an accident. The second is to reduce the exposure to the environment within which the accidents take place. While most of the efforts are focused on the first solution (with little success), the second solution is often ignored. However, it may be time to ask, how many of the operations flown are true emergencies? Should HEMS operations be used to transfer stable patients, and if so, under what conditions should these be considered? How are go/no go decisions made, and by who? At what point should a transport be turned down?

Taken altogether, these recommendations may help to alleviate some of the problems facing HEMS today. For as the data pointed out, we are not facing one problem, but numerous issues, all of which must be addressed. In particular, we need to go beyond the aircrew and study the system in which they operate. This should include a two pronged approach: 1) We must understand the culture within which the aircrews operate. This should include not only the supervisory and organizational issues, but aircraft environment such as protective clothing, helmets, instrument tie downs, etc… and 2) a detailed analysis of current regulations in order to understand how regulatory practices interact with HEMS operations. Only by doing this can we hope to fix the system.

REFERENCES

Association of Air Medical Services.(2001). AAMS' response to HCFA with regard to patient privacy concerns. http://www.aams.org/hipaa.html

Blumen, I.J. (2002). A safety review and risk assessment in air medical transport; *supplement to the Air Medical Physician Handbook*. Copyright 2002, Air Medical Physician Association.

Cowley, R.A. (1976). The resuscitation and stabilization of major multiple trauma patients in a trauma center environment. *Clinical Medicine*, *83*, 14-22.

Frazer, R. (1999). Air medical accidents: A 20 year search for information. *AirMed Jouna*l, September/October, 34-39.

National Academy of Sciences, National Research Council. Accidental Death and Disability, the neglected disease of modern society. Washington, D.C., US Government Printing Office, 1966, pp 5-6, 12, 15-16, 21, 35.

Reason, J. (1990). *Human Error*. New York, University Press.

Thomas, F (1988). The development of of the nation's oldest operating civilian hospital-sponsored aero medical helicopter service. *Aviation, Space and Environmental Medicine*, 59, 567.

Wiegmann, D. & Shappell, S. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Burlington, VT: Ashgate.

HUMAN ERROR AND COMMERCIAL AVIATION ACCIDENTS: A COMPREHENSIVE, FINE-GRAINED ANALYSIS USING HFACS

Scott Shappell, Ph.D.
Clemson University
Clemson, SC

Albert Boquet, Ph.D.
Embry-Riddle Aeronautical University
Daytona Beach, FL

Cristy Detwiler, B.A
Kali Holcomb, A.A.
Carla Hackworth, Ph.D.
Cristina Bates
Civil Aerospace Medical Institute
Oklahoma City, OK

Douglas Wiegmann, Ph.D.
Mayo Clinic
Rochester, MN

The Human Factors Analysis and Classification System (HFACS) is a theoretically based tool for investigating and analyzing human error. The aim of this study was to extend previous examinations of aviation accidents to include specific aircrew, environmental, supervisory, and organizational factors associated with 14 CFR Part 121 (Air Carrier) and 14 CFR Part 135 (Commuter) accidents using HFACS. The majority of causal factors were attributed to the aircrew and the environment with decidedly fewer associated with supervisory and organizational causes. Recommendations were made based on the HFACS findings presented.

INTRODUCTION

While commercial[1] aviation accident rates have reached unprecedented levels of safety, little, if any, improvement has been realized over the last decade for either the air carrier or commuter/air taxi industry (Figure 1). Indeed, some have even suggested that the current accident rate is as good as it gets – or is it?

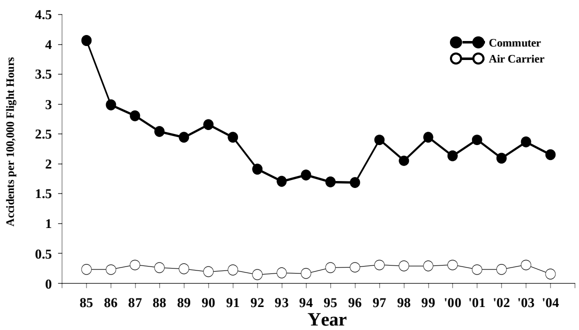


Figure 1. Air carrier and commuter/air taxi accident rates since 1985 (Source: NTSB).

The challenge for the Federal Aviation Administration (FAA) and other civil aviation safety organizations is to improve an already very safe industry. The question is where to start when most of the "low hanging fruit" (e.g., improved powerplant and airframe technology, advanced avionics, and the introduction of automation) have been "picked."

It is typically reported that somewhere between 60-80% of aviation accidents are due, at least in part, to human error (Shappell & Wiegmann, 1996). That being said, it may be surprising that with few exceptions (e.g., Billings & Reynard, 1984; Gaur, 2005; Li, Baker, Grabowski, & Rebok, 2001; Shappell & Wiegmann, 2003a, 2003b; Wiegmann & Shappell, 2003) most studies to date have focused on situational factors or pilot demographics, rather than the underlying human error causes of accidents.

Judging from current accident rates, situational and pilot demographic data alone have provided little in the way of preventing accidents, apart from identifying target populations for the dissemination of safety information. Given the multi-factorial nature of accidents (Baker, 1995), it may make more sense to examine these variables within the context of what we know about human error and accident causation.

HFACS

It is generally accepted that aviation accidents are typically the result of a chain of events that often culminate with the unsafe acts of operators (aircrew). The aviation industry is not alone in this belief, as the safety community has embraced a sequential theory of accident investigation since Heinrich first published his axioms of industrial safety in 1931 (Heinrich, Peterson, & Roos, 1931). When Reason published his "Swiss cheese" model of human error in 1990, the aviation community began a systematic examination of human error.

Drawing upon Reason's (1990) concept of latent and active failures, HFACS describes human error at each of four levels: 1) the unsafe acts of operators (e.g., aircrew, maintainers, air traffic controllers), 2) preconditions for unsafe acts, 3) unsafe supervision

[1] The FAA distinguishes between two types of commercial operations: those occurring under 14 CFR Part 121 – Air Carrier Operations and those occurring under CFR Part 135 commuter/air taxi operations.

(i.e., middle-management), and 4) organizational influences.[2]

Unsafe Acts of Operators

The unsafe acts of operators (aircrew) can be loosely classified into one of two categories: errors and violations (Reason, 1990). While both are common within most settings, they differ markedly when the rules and regulations of an organization are considered. Errors represent authorized behavior that fails to meet the desired outcome. Whereas, violations refer to the willful disregard of the rules and regulations. It is within these two overarching categories that HFACS describes three types of errors [decision (DE), skill-based (SBE), and perceptual (PE)] and two types of violations (V, routine and exceptional).

Preconditions for Unsafe Acts

Simply focusing on unsafe acts, however, is not enough. Investigators must dig more deeply into the preconditions for unsafe acts. Within HFACS, three major subdivisions are described: 1) condition of the operator, 2) personnel factors, and 3) environmental factors.

Unsafe Supervision

Clearly, aircrews are responsible for their actions and, as such, must be held accountable. However, in some instances, they are the unwitting inheritors of latent failures attributable to those who supervise them. To account for these, the overarching category of unsafe supervision was created within which four categories (inadequate supervision, planned inappropriate operations, failed to correct known problems, and supervisory violations) are included.

Organizational Influences

Where decisions and practices by front-line supervisors and middle-management can adversely impact aircrew performance, fallible decisions of upper-level management may directly affect supervisors and the personnel they manage. The HFACS framework describes three latent organizational failures: 1) resource management, 2) organizational climate, and 3) operational processes.

PURPOSE

The goal of the present study was twofold: 1) to extend our previous HFACS analyses beyond military and general aviation (GA) to include a comprehensive analysis of commercial aviation; and 2) to combine the power of a theoretically derived human error framework (i.e., HFACS) with traditional situational and demographic data from the accident records.

METHOD

Data

Commercial aviation accident data (i.e., 14 CFR Part 121 – air carrier; 14 CFR Part 135 – commuter) from calendar years 1990-2002 were obtained from databases maintained by the National Transportation Safety Board (NTSB) and the FAA's National Aviation Safety Data Analysis Center (NASDAC).

Eliminated from consideration were accidents that were classified as having "undetermined causes," and those that were attributed to sabotage, suicide, or criminal activity (e.g., stolen aircraft). The data were culled further to include accidents that involved aircrew or supervisory error. Of the remaining 1,020 accidents, 181 involved air carrier aircraft and 839 involved commuter aircraft.

Causal Factor Analysis Using HFACS

Six pilots served as subject matter experts (SMEs). All were certified flight instructors with a minimum of 1,000 flight hours at the time they were recruited.

Each pilot was provided roughly 16 hours of instruction on the HFACS framework. After training, the pilots were randomly assigned accidents such that at least two separate pilots independently analyzed each accident.

Using narrative and tabular data obtained from both the NTSB and the FAA NASDAC, the pilots classified each aircrew or supervisory causal factor identified by the NTSB using the HFACS framework. Where disagreements existed, the corresponding pilots were instructed to reconcile their differences. Overall, pilots agreed more than 85% of the time.

RESULTS

A summary of the HFACS analyses of commercial aviation accidents can be found in Table 1. The majority of human causal factors identified involved aircrew and their environment (i.e., *unsafe acts of operators* and *preconditions for unsafe acts*) rather than supervisory or organizational factors. Nevertheless, when organizational influences were observed they typically involved *operational processes* such as inadequate or non-existent procedures, directives, standards, and/or requirements or in the case of commuter operations, inadequate surveillance of operations. Unsafe supervision on the other hand, typically involved *inadequate supervision* in general or the failure to provide adequate training.

As anticipated, a large number of *environmental conditions* were identified within the commercial aviation database, particularly those associated with

[2] A complete description of all 19 HFACS causal categories is available elsewhere (see Wiegmann & Shappell, 2003)

aspects of the *physical environment* like weather and lighting. However they were not uniformly distributed across air carrier and commuter operations, as considerably more issues associated with the *physical environment* were observed during commuter (63%) than air carrier operations (37%). In contrast, the accident record revealed surprisingly few problems associated with the technological environment. Preconditions associated with aircrew were also frequently observed within the accident record. For instance, *crew resource management (CRM)* failures were identified in nearly one out of every five air carrier accidents examined. Even more interesting, the nature of the CRM failure differed between the two commercial operations. That is, while over 60% of the CRM failures associated with air carrier accidents involved "inflight" CRM failures (e.g., inflight crew coordination, communication, monitoring of activities, etc.), over 80% of the CRM failures observed during commuter operations involved "preflight" activities (such as planning and briefing).

As seen in other aviation operations (Shappell & Wiegmann, 1995, 1997, 2003a, 2003b, 2004; Wiegmann & Shappell, 1997, 2001a, 2001b, 2003) the majority of commercial aviation accident causal factors were found at the unsafe act level. Indeed, just over half of the accidents were associated with at least one SBE, followed by DEs (36.7%) and Vs of the rules and regulations (23.1%).

Similar to other civil aviation accident data (Shappell & Wiegmann, 2003a, 2003b, 2004; Wiegmann & Shappell, 2003), there was little variation in the distribution of unsafe acts committed annually by aircrew flying either air carrier or commuter operations (Figure 2A & 2B). When accidents occurred in either type of commercial operation, they were typically associated with more SBEs followed by DEs, Vs, and PEs respectively. This was true even though the air carrier data had to be averaged over 3 or 4 year blocks due to the small number of accidents in the database (Figure 2A). Moreover, with the exception of the violations category which has shown a slight increase since the 1993-1995 time frame, the annualized data were relatively flat suggesting that there has been little impact on any specific type of human error over the last 13 years.

Table 1. Frequency and percentage of accidents associated with each HFACS causal category by type of operation.

HFACS Category	Air Carrier	Commuter	Total
Organizational Influences	**N (%)**	**N (%)**	**N (%)**
Resource Management	4 (2.2)	0 (0.0)	4 (0.4)
Organizational Climate	0 (0.0)	4 (0.5)	4 (0.4)
Operational Process	21 (11.6)	29 (3.5)	50 (4.9)
Unsafe Supervision			
Inadequate Supervision	15 (8.3)	21 (2.5)	36 (3.5)
Planned Inappropriate Operations	3 (1.7)	5 (0.6)	8 (0.8)
Failed to Correct Known Problems	0 (0.0)	0 (0.0)	0 (0.0)
Supervisory Violations	0 (0.0)	2 (0.2)	2 (0.2)
Preconditions of Unsafe Acts			
Environmental Conditions			
Technological Environment	11 (6.1)	4 (0.5)	15 (1.5)
Physical Environment	67 (37.0)	525 (62.6)	592 (58.0)
Conditions of the Operator			
Adverse Mental States	6 (3.3)	60 (7.2)	66 (6.5)
Adverse Physiological States	6 (3.3)	18 (2.1)	24 (2.4)
Physical/Mental Limitations	6 (3.3)	39 (4.6)	45 (4.4)
Personnel Factors			
Crew Resource Management	34 (18.8)	75 (8.9)	109 (10.7)
Personal Readiness	0 (0.0)	3 (0.4)	3 (0.3)
Unsafe Acts of the Operator			
Skill-based Errors	77 (42.5)	499 (59.5)	576 (56.5)
Decision Errors	71 (39.2)	303 (36.1)	374 (36.7)
Perceptual Errors	10 (5.5)	56 (6.7)	66 (6.5)
Violations	31 (17.1)	205 (24.4)	236 (23.1)

Note: Numbers in the table involve at least one instance of an HFACS category. For example 77 of the 181 air carrier accidents (77/181 or 42.5%) were associated with at least one skill-based error. Because accidents are generally associated with more than one causal factor, the percentages in the table do not add up to 100%.

14 CFR Part 135 - Commuter Operations

Because of the relatively small number of air carrier accidents in the database related to aircrew/supervisory error, additional fine-grained analyses of those data were not possible. However, more detailed analyses were conducted for commuter operations.

Visual Conditions. Given the relatively large percentage of accidents associated with physical conditions, in particular those associated with prevailing weather conditions and lighting, it seemed reasonable to begin with these two environmental causal factors. As can be seen in Figure 3A, just over 70% of the accidents occurred during visual meteorological conditions (VMC). Likewise, roughly 70% of the accidents occurred in broad daylight (Figure 3B).

In order to capitalize on the threat posed by both environmental causal factors, the two were combined to create a new variable. Specifically, two levels of visual conditions were created: 1) clear visual conditions which included accidents that occurred during VMC and daylight conditions, and 2) impoverished visual conditions that included accidents occurring during instrument meteorological conditions (IMC) or at twilight/night.

Unlike the results seen with weather and lighting conditions alone, the recombination of visual factors showed that the percentage of accidents occurring in clear visual conditions was only marginally higher than that occurring in visually impoverished conditions (Figure 3C). It would appear that while weather and lighting conditions are important factors in aviation, their impact is potentially magnified when a pilot's ability to see outside the aircraft is taken into consideration.

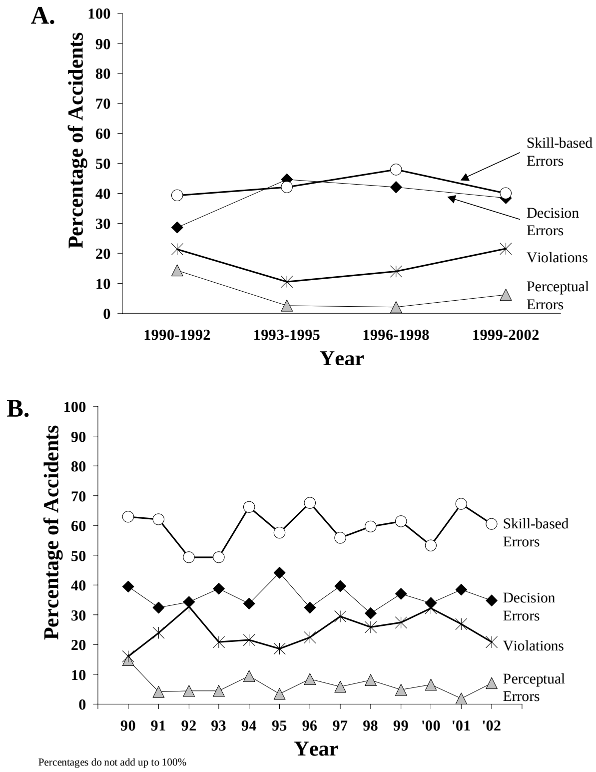


Figure 2. Percentage of unsafe acts committed by aircrew during air carrier (Panel A) and commuter (Panel B) operations by year.

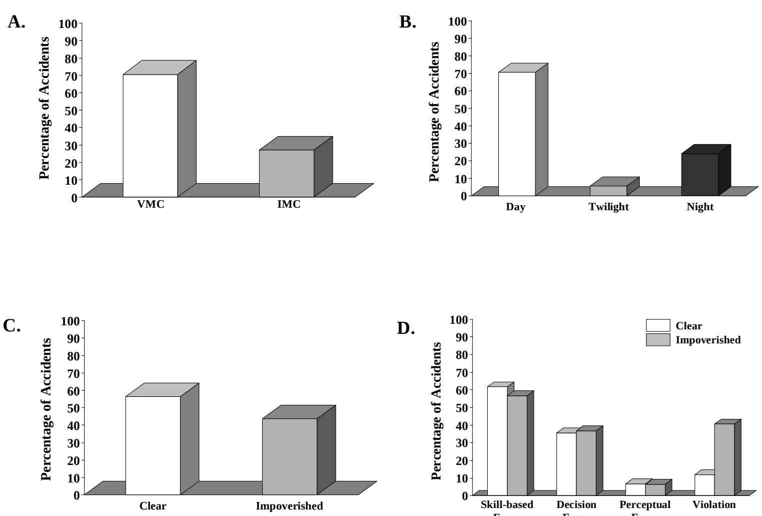


Figure 3. Percentage of commuter accidents by weather conditions (Panel A), lighting conditions (Panel B), visual conditions (Panel C) and visual conditions by unsafe acts (Panel D).

Naturally, one would expect the pattern of human error to be different during accidents in clear versus visually impoverished conditions. Indeed, when visual conditions were compared across the unsafe acts of aircrew (Figure 3D), an interesting pattern of human error emerged. While SBEs were the most common error form observed during accidents in clear and impoverished conditions, Vs were five times more likely to be attributed to accidents in visually impoverished conditions (X^2 = 92,322, p<.001; odds ratio = 5.077).

Upon closer examination, intentional flight into IMC while operating under visual flight rules (i.e., VFR flight into IMC) accounted for nearly 1/3 of the Vs observed during impoverished visual conditions. In addition, the failure to adhere to procedures/directives (V), poor in-flight planning/decision making (DE), the loss of control in-flight (SBE), and the failure to maintain sufficient airspeed (SBE) all were commonly cited as causes during accidents in visually impoverished conditions.

The failure to adhere to procedures/directives (V) was also frequently seen among accidents in clear conditions as was poor in-flight planning/decision-making (DE). However, unlike impoverished visual conditions, commuter accidents occurring in the clear were often associated with the selection of unsuitable terrain (DE) and the inability to compensate for winds (SBE).

Injury Severity. Previous investigations of GA accidents have shown distinct differences in the pattern of human error associated with fatal and non-fatal aviation accidents (Shappell & Wiegmann, 2003a, 2003b; Wiegmann & Shappell, 2003). A similar examination of commuter accidents revealed that roughly 30% of all commuter accidents resulted in at least one fatality.

As with the findings regarding visual conditions, SBEs were associated with the majority of fatal and non-fatal accidents followed by DEs, Vs, and PEs. Of note however, Vs were more than three times as likely to be associated with fatal accidents (X^2 = 48.239, p<.001; odds ratio = 3.145).

Upon closer examination, it appears that causal factors such as intentional VFR flight into IMC (V), poor in-flight planning/decision making (DE), and control of the aircraft and airspeed (SBE) were the most frequently cited aircrew errors associated with fatal accidents. In contrast, non-fatal accidents appear to be more closely associated with the failure to compensate for winds (SBE), loss of directional control on the ground (SBE), selection of unsuitable terrain (DE), poor in-flight planning/decision-making (DE), and the failure to follow procedures/directives (V).

Given the similarity in the pattern of human errors associated with visual conditions and injury severity (fatal vs. non-fatal), it made sense to examine the combination of the two variables. The largest percentage of fatal commuter accidents occurred in visually impoverished conditions. In contrast, when the accident occurred in clear visual conditions, a much smaller percentage resulted in fatalities. Indeed, commuter accidents were over four times more likely to result in fatalities if they occurred in visually impoverished conditions (X^2 = 83.978, p<.001; odds ratio = 4.256).

Fully one half of the fatal accidents occurring in visually impoverished conditions involved at least one V – often intentional VFR flight into IMC. Not surprisingly, given the environmental conditions at the time, poor in-flight planning (DE) was also commonly cited among this subset of the data.

DISCUSSION

Generally speaking, nearly 70% of the "commercial" aviation accidents occurring between 1990 and 2002 were associated with some manner of aircrew or supervisory error. However, the percentage varied slightly when air carrier (45%) and commuter (75%) aviation accidents were considered separately. This finding is consistent with results reported elsewhere (Li, Baker, Grabowski, & Rebok, 2001).

Organizational Influences and Unsafe Supervision

Consistent with previous work (Wiegmann & Shappell, 2001a), comparatively few commercial aviation accidents were associated with organizational and/or supervisory causal factors - particularly within the commuter aviation industry. In spite of this, a relatively large proportion of accidents involved issues related to *operational processes*. Causal factors associated with the remaining HFACS organizational causal categories, *resource management* and *organizational climate*, were rarely observed in the data.

A closer inspection revealed that the particular type of *operational process* cited appeared to be dependent on the type of operation involved. Namely, air carrier accidents were typically associated with the manner in which procedures or directives were communicated assuming they existed at all. In contrast, commuter accidents were more often associated with a lack of organizational oversight. Exactly why this difference might exist requires a more in-depth investigation than what was performed here.

Preconditions for Unsafe Acts (Aircrew)

With a couple of notable exceptions causal categories within the preconditions for unsafe acts were also lightly populated. One of those exceptions was the large proportion of accidents (particularly among commuter aviation) influenced by prevailing weather conditions and reduced visibility. This was not particularly surprising since studies like the one conducted by Baker, Lamb, Li, and Dodd (1993) reported similar results in their examination of commuter accidents between 1983 and 1988.

While previous efforts suggested that factors associated with the *physical environment* and *CRM* would be identified among the commercial data, it was surprising that other areas, in particular the *condition of the operator (aircrew)*, were not identified in the accident record more often. The exception involved commuter aviation accidents, where a number of *adverse mental states* (64 out of 839 accidents or 7.2%) and *physical/mental limitations* (43 out of 839 or 4.6%) were observed.

In some ways the fact that many commuter aviation operations are single-piloted may explain why *adverse mental states* played a more prominent role among these accidents. For instance, without a second set of eyes in the cockpit any distraction would likely be exacerbated and distract the pilot from the task at hand – flying the aircraft.

Perhaps more disconcerting than the issue of attention was the large number of commuter aviation accidents associated with the pilot's lack of experience – something rarely seen among the air carrier accidents examined. Whether this represents a lack of flight hours or merely inexperience with a particular operational setting or aircraft remains to be determined. Still, flight hours alone may not be sufficient to overcome the lack of experience observed here. After all, flying straight and level in VMC will not prepare a pilot for the complexities of instrument flight or the dangers of flying in other potentially hazardous environments.

Unsafe Acts of Operators (Aircrew)

As with our previous efforts involving civil and military aviation (Wiegmann & Shappell, 1997, 1999, 2001a, 2001b), SBEs were the most prevalent form of aircrew error among the commercial aviation accidents examined. Particularly widespread were technique errors associated with handling or controlling the aircraft. More important, when the commercial data reported here were combined with our previous investigations of GA accidents (Wiegmann, Shappell, Boquet, Detwiler, Holcomb, Faaborg, in press; Detwiler, Hackworth, Holcomb, Boquet, Pfleiderer, Wiegmann, and Shappell, in review) an interesting finding emerged. It appears that the percentage of SBEs associated with accidents increases systematically as one moves from air carrier (43%) to commuter (60%) to GA (73%) operations.

At first glance, this would appear to suggest that pilot skill and proficiency is best among the air carrier industry and becomes progressively more suspect within commuter and GA. Recall that SBEs, by definition, occur during the execution of routine events (Reason, 1990; Rasmussen, 1982). Furthermore, once a particular skill is developed, it must be maintained through repetition and experience. That being said, most people would agree that GA pilots fly less and participate in fewer recurrent training sessions than their commercial counterparts. It stands to reason that their proficiency would be less than their commercial counterparts and may explain why SBEs are more prevalent among GA accidents.

DEs were observed in roughly four out of every ten commercial aviation accidents while Vs and PEs were observed in 23% and 7% of the accidents, respectively. Some have even argued that DEs and Vs are of the same ilk (i.e., both involve decisions by aircrew that go awry) and should actually be combined in the HFACS framework. If this were true, the combined causal category of DE/V would be roughly equivalent to that seen with SBEs.

Scenario-based training, in-flight planning aids, and education may improve pilot decision-making; however, these approaches have been largely ineffective in stemming Vs. Instead, enforcing current standards and increasing accountability in the cockpit may be the only effective means to reduce violations of the rules – a tactic that is often difficult to employ in civil aviation. As a result, the FAA and the commercial aviation industry may have to look to other avenues to reduce Vs such as the use of flight simulators that can demonstrate the hazards associated with violating the rules (Knecht, Harris, & Shappell, 2003).

REFERENCES

Baker, S. (1995). Putting "human error" into perspective. *Aviation, Space, and Environmental Medicine, 66*, 521.

Baker, S., Lamb, M., Li, G., & Dodd, R. (1993). Human factors in crashes of commuter airplanes. *Aviation, Space, and Environmental Medicine, 64*, 63-68.

Billings, C. & Reynard, W. (1984). Human factors in aircraft incidents: Results of a 7-year study. *Aviation, Space, and Environmental Medicine, 55*, 960-965.

Detwiler, C., Hackworth, C., Holcomb, K., Boquet, A., Pfleiderer, E., Wiegmann, D., & Shappell, S. (in review). Beneath the tip of the iceberg: A human factors analysis of general aviation accidents in Alaska verses the rest of the United States. Federal Aviation Administration, Office of Aerospace Medicine Technical Report No. DOT/FAA/AM-05/xx. Office of Aerospace Medicine: Washington, DC.

Gaur, D. (2005). Human factors analysis and classification system applied to civil aircraft accidents in India. *Aviation, Space, and Environmental Medicine, 76*, 501-505.

Heinrich, H., Peterson, D., & Roos, N. (1931). *Industrial accident prevention: A safety management approach* (1st ed.). New York, NY: McGraw-Hill.

Knecht, W., Harris, H., & Shappell, S. (2003). Effects of visibility, cloud ceiling and financial incentive on general aviation voluntary takeoff into adverse weather. *Proceedings of the 12th International Symposium on Aviation Psychology*, 669-73.

Li, G., Baker, S., Grabowski., J., & Rebok, G. (2001). Factors associated with pilot error in aviation crashes. *Aviation, Space, and Environmental Medicine*, *72*, 52-58.

Rasmussen, J. (1982). Human errors: A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4*, 311-33.

Reason, J. (1990). Human Error. New York: Cambridge University Press.

Shappell, S. & Wiegmann, D. (1995). Controlled flight into terrain: The utility of an information processing approach to mishap causal factors. *Proceedings of the Eighth International Symposium on Aviation Psychology*, The Ohio State University, 1300-06.

Shappell, S. & Wiegmann, D. (1996). U.S. naval aviation mishaps 1977-92: Differences between single- and dual-piloted aircraft. *Aviation, Space, and Environmental Medicine, 67*(1), 65-69.

Shappell, S. & Wiegmann, D. (1997). Why would and experienced aviator fly a perfectly good aircraft into the ground? *Proceedings of the Ninth International Symposium on Aviation Psychology*, The Ohio State University, 26-32.

Shappell, S. & Wiegmann, D. (2003a). A human error analysis of general aviation controlled flight into terrain (CFIT) accidents occurring between 1990-1998. Federal Aviation Administration, Office of Aerospace Medicine Technical Report No. DOT/FAA/AM-03/4. Office of Aerospace Medicine: Washington, DC.

Shappell, S. & Wiegmann, D. (2003b). Reshaping the way we look at general aviation accidents using the human factors analysis and classification system. *Proceedings of the Twelfth International Symposium on Aviation Psychology*, Dayton, Ohio.

Shappell, S. & Wiegmann, D. (2004). HFACS Analysis of Military and Civilian Aviation Accidents: A North American Comparison. *Proceedings of the Annual Meeting of the International Society of Air Safety Investigators*, Gold Coast, Australia.

Wiegmann, D. & Shappell, S. (1997). Human factors analysis of post-accident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology, 7*(1), 67-81.

Wiegmann, D. & Shappell, S. (1999). Human error and crew resource management failures in naval aviation mishaps: A review of U.S. Naval Safety Center data, 1990-96. *Aviation, Space, and Environmental Medicine, 70*(12), 1147-51.

Wiegmann, D. & Shappell, S. (2001a). Human error analysis of commercial aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Aviation, Space and Environmental Medicine, 72*, 1006-16.

Wiegmann, D. & Shappell, S. (2001b). Human error perspectives in aviation. *International Journal of Aviation Psychology, 11*(4), 341-57.

Wiegmann, D. & Shappell, S. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Burlington, VT: Ashgate.

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

SYMBOL SET DISCRIMINABILITY METRICS: EXTENDING DISCRIMINATION MODELS FOR SIZE AND POSITION INVARIANCE

Albert J. Ahumada, Jr., Ph.D., Maite Trujillo San Martin, Ph.D., Jennifer Gille, Ph.D.
NASA Ames Research Center, Moffett Field, California

If flight display symbols are to be safely recognized by pilots, they need to be easily discriminated from each other. A study by Michael Zuschlag, DOT Volpe Center, assessed the recognizability of a proposed traffic symbol set. Predictions for the study results were generated by a standard image discrimination model. This model predicts that any difference whatever in the two images presented to it contributes to discriminability, while the observers appeared to categorize somewhat independently of size and position. An image discrimination model was developed that included both size compensation and position compensation. We applied this model to seven of the symbol pairs that lead to the most errors in the Volpe experiment. The predictions of experimental results by the model were improved. The model takes as input the luminance values for the pixels of two symbol images, the effective viewing distance, and gives as output the discriminability in just-noticeable-differences (d'), the size reduction of the larger symbol, and the x and y offset in pixels needed to minimize the discriminability.

INTRODUCTION

The goal of this project is to provide tools that can be used to evaluate the discriminability of symbols using extensions of visual discrimination models (Ahumada and Beard, 1998; Beard, Jones, Chacon, and Ahumada, 2005). Discriminability is only one component of the property Yey and Chandra (2004) call distinctiveness, the degree to which the symbol by itself can be identified. The visual discrimination models do not have a theory of feature learning or feature extraction or attention or memory effects. Bruner (1973) gives a good overview of these higher level processes that can affect symbol categorization. We are concerned here with symbol discriminability that only depends on low level visual processes.

Initially we planned to provide a model similar to that reported by Watson and Ahumada (2004; 2005). That model predicted the accuracy of letter identification in an acuity task as a function of optical distortions of the letters. We decided that in actual usage, the symbols would not be used with equal frequency and that actual performance correct was not as important as the possibility of potential errors. We decided to provide a tool that could be used to measure the discriminability of pairs of stimuli. All pairs in a set of potential symbols would need to be compared to ensure discriminability, but discriminability would not ensure accurate categorization. For example in the color domain, it is well known that many colors are discriminable from each other but that relatively few categories of colors can be accurately reported by naïve observers (Miller, 1956; Garner, 1962).

APPROACH

We began by looking at the data from the Volpe experiment, whose methodology and results were summarized by Zuschlag (2004).

"Methodology: The study is a descriptive psychophysical experiment. Ten pilots were recruited from a local airport. All had normal color vision and adequate visual acuity. The 19 symbols in the symbol set were presented

one at a time on a bench-mounted aviation multifunction display (MFD) for 250 ms. The MFD was illuminated with approximately 94 kLx of light using a spotlight to simulate sun-shaft illumination. For each trial, each participant was shown a symbol in isolation and asked to select the perceived symbol from a matrix of 19 possible symbols presented on a laptop equipped with a touch screen. Error rates and reaction time were recorded.
Results: When viewed at a distance and angle approximating that found in a general aviation cockpit, most symbols were correctly recognized at least 92% of the time. The exception was symbols intended to indicate a selected state; these were correctly recognized as low as 83% of the time. In the proposed symbol set, a selected state was indicated by outlining the symbol. The data suggest that this convention increases the likelihood that participants will confuse symbols indicating non-proximal traffic (represented by a hollow symbol) with symbols indicating proximal traffic (represented by a solid symbol)."

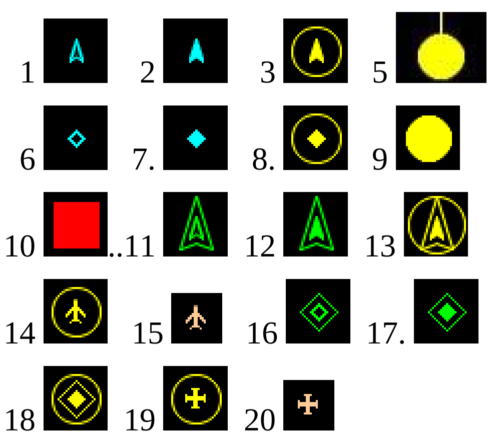


Figure 1. The Volpe experiment symbols.

Some of the results that appeared to us are the following:

1) There are significant order effects that are confounded with observer effects because the design was not balanced. Observer 5 made more than twice as many errors as anyone else, but he was the only one to begin with the farthest distance.

2) There are significant rotation effects. They only seemed to occur for the selected directional symbols. We will not be able to tell if these are perceptual or cognitive (categorical) without knowing what the actual images were (the rotation may have changed the outline). When Image 13 was in rotation 1 (we do not know what the positions mean), it was almost always called 13, but in other orientations it was often called 3 (see Figure 2).

Figure 2. Symbols 13 and 3.

3) The errors are not symmetric. A line version is called a filled version, but not the reverse. At the 88 in. distance, if wrong, Image 1 was most likely to be called 2, but Image 2 was most likely to be called 12 (see Figure 3), an indication of size invariance. At that distance, Image 11 was more likely to be called 12 than 11 (see Figure 3).

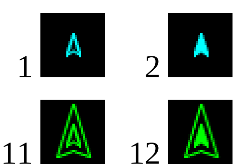


Figure 3. Image pairs 1 and 2 and 11 and 12.

The confusion of Image 13 being called 3 is also consistent with the filling-in principle and size invariance (see Figure 2).

Such principles imply that simple image difference models can not predict the actual pattern of responses. Another possible cause of response asymmetries is that the observer scanned through the responses sequentially, stopping when a match occurred without considering other possibilities.

4) Symbols 15 and 20 were mainly confused with each other, with a strong bias for responding that the plane (15) was present rather than the cross (20). At the 88 in. distance, they were actually more likely to say that the plane was present when the cross was. At that distance, the 2x2 confusion matrix has a d' of 0.6, while the image discrimination model predicts a d' of 1.1 (under various assumptions about the presentation conditions that we know are wrong). This result suggests the model can do a fair job of predicting the observed discriminability, since the observers vary in sensitivity by at least a factor of three.

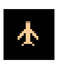
15  20 

Figure 4. Symbol pair 15 and 20.

From this analysis of the data from the Volpe experiment, we decided that a key problem with the existing discrimination models is that they do no compensate for pattern recognition transformational invariances that are naturally made by human observers. In the intended application of the symbols on a spatial map, the observers obviously must report that a symbol is the same when it is translated to a new position. Symbol size could be used as a cue, but, since the size would have to be anchored and is subject to strong context effects, we assume that symbols that are similar when one is magnified relative to the other will not be reliably discriminated. This transformation was suggested by the confusions between symbols 3 and 13 in the Volpe study despite the large discriminability predicted by the simple image discrimination model.

Size compensation is implemented by a frequency domain image shrinking algorithm (Watson, 1986). The current version only shrinks even sized images to even sized images, and the proportion lower range is an input parameter set to 0.5. This would have resulted in only 8 values of shrinkage for the 32 by 32 pixel symbols, so we pixel replicated the images by a factor of four so that 32 shrinkage levels were assessed. Higher resolution can be obtained by increased pixel replication of the images. The first pixel of the image must be the background level for the image because it is used to extend the smaller image so that all images are the same size before and after the size adjustment. Position compensation invariance is implemented by cross correlating the visual contrast images as a function of spatial offset by frequency domain filtering of one image by other. The pixel replication by a factor of four results in the position search being done to 0.25 pixel accuracy.

We have implemented Matlab routines to compute all the following steps:

1) Shrinking and padding of an image.

2) Conversion to visible contrast images.

5) Computation of the minimum visible difference position offset and the actual minimum visible difference there.

The code is available at (http://vision.arc.nasa.gov/personnel/al/code/index.htm).

RESULTS

Figure 5 shows some preliminary results from the model with size and position compensation. These calculations were done for full contrast images. Without size compensation, the model prediction of difference between symbols 3 and 13 was a d' of 4.8. With size compensation the predicted d' is lowered to 3.5. These d' values are slightly larger than those we previously reported without size and position compensation because the contrast sensitivity parameter was increased so that the best predicted detection performance for an observer would be 0 dBB (Watson, 2000).

The most interesting result was that the model now predicts that surrounding the plane and cross with a circle improves the discriminability (though not as much as observed). The plain image discrimination model with no masking predicts no effect of the circle. Adding masking to the model causes the model to predict that the circles will make the difference even more difficult to detect. Adding the translation invariance to the model allows the cross to be moved down to a better fitting position than is possible when the circle is present and stabilizes the position. The model thus pointed out that having the circle present allows one to see more clearly the asymmetric nature of the plane.

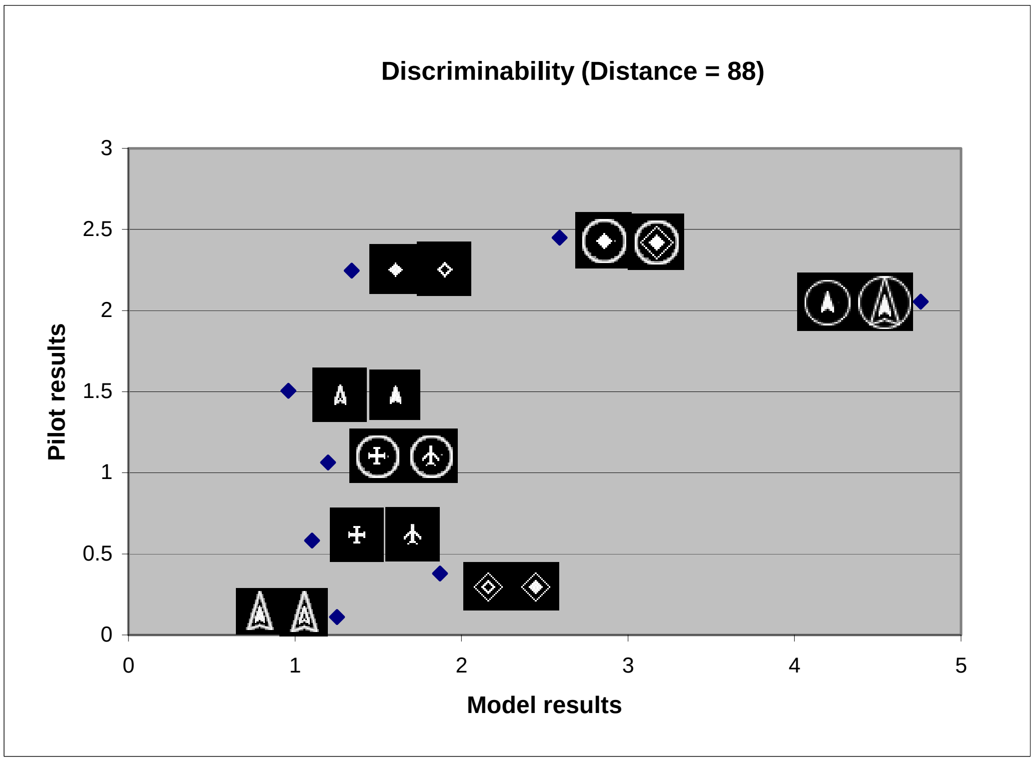


Figure 5. Predictions of the model with size and position compensation.

DISCUSSION

The implementation of the size compensation brought up two new issues that we had not considered before. One is that size compensation could be done separately in the x and y directions, so that the model would also predict confusions when a symbol is a taller or shorter version of another symbol. This feature would probably have helped the letter recognition predictions of Watson and Ahumada (2004). Another discovery is that when two images are different and about the same size there can be a slight advantage in having either of them made slightly smaller. So far, this effect has been small enough to be neglected.

There are two main issues that we have not dealt with. One is color and the other is orientation. In the Volpe experiment all the symbols were of a single color and there were essentially no confusions between differently colored symbols. For such symbol sets, the discriminability could be easily handled by adding color as a two more dimensions (essentially two more pixels). For symbol sets with multi-colored symbols, we would need to add two more images to each image in the manner of the detection model of Wuerger, Watson, and Ahumada (2002), together with the masking model of Ahumada and Krebs (2001)

As mentioned above, the Volpe study suggests that there were differences in the confusions as a function of orientation. If this result is not an artifact of rendering or of the fact that only one orientation was available in the response set, it would indicate that a simple model that extracted the orientation and the pattern from a "standard" orientation will not work. At this point the discriminability as a function of orientation can only be evaluated by brute force.

REFERENCES

Ahumada, A. J. Jr., Beard, B. L. (1998) A simple vision model for inhomogeneous image quality assessment, in SID Digest of Technical Papers, ed. J. Morreale, vol. 29, Paper 40.1, Santa Ana, CA.

Ahumada, A. J. Jr., and Krebs, W. K. (2001) Masking in color images, in B. E. Rogowitz and T. N. Pappas, Eds., Human Vision and Electronic Imaging VI, SPIE Proc. Vol. 4299, pp. 187-194.

Beard, B. L., Jones, K. M., Chacon, C., and Ahumada, A. J., Jr. (2005) Detection of blurred cracks: A step towards an empirical vision standard, Final Report for FAA Agreement DTFA-2045.

Bruner, J. S. (1973) Beyond the information given: Studies in the psychology of knowing. Oxford, UK: W. W. Norton.

Garner, W. R. (1962) Uncertainty and Structure as Psychological Concepts, New York: John Wiley and Sons, Inc.

Miller, G. (1956) The magical number seven, plus or minus two. Psychol. Rev. 63, 81-97.

Zuschlag, M (2004) Assessment of proposed traffic symbol set. FAA Human Factors Research and Engineering Division FY2004 Report, p. 44. https://www.hf.faa.gov/docs/508/docs/2004Report.pdf.

Watson, A. B. (1986) Ideal shrinking and expansion of discrete sequences, NASA Technical Memorandum 88202.

Watson, A. B., and Ahumada, A. J. (2004) Human optical image quality and the Spatial Standard Observer, OSA Fall Vision Meeting,

Watson, A. B., and Ahumada, A. J. (2005) Predicting acuity from aberrations with the spatial standard observer, Investigative Ophthalmology and Visual Science 46, E-Abstract 3614.

Wuerger, S., Watson, A. B., and Ahumada, A. J. Jr. (2002) Toward a standard observer for spatio-chromatic detection. SPIE Proc. Vol. 4662, 159-172.

Yeh, M., and Chandra, D. (2004) Issues in symbol design for electronic displays of navigation information. Proceedings of the 23rd DASC Conference, 24-28 October, Salt Lake City, UT.

ACKNOWLEDGEMENTS

This work was partly supported by the HMP Project of NASA's Airspace Systems Program. Dr. Trujillo is a National Research Council Associate and Dr. Gille was supported through the University Affiliated Research Center at NASA Ames Research Park.

TRANSFER OF TRAINING EFFECTIVENESS OF A FLIGHT TRAINING DEVICE (FTD)

Henry L. Taylor, Donald A. Talleur, Tom W. Emanuel, Jr., and Esa M. Rantanen,
Institute of Aviation, University of Illinois at Urbana-Champaign
Savoy, Illinois

A transfer of training research design was used to measure the effectiveness of a flight training device (FTD) and to determine the point at which additional training in a FTD was no longer effective. The dependent measures were number of trials to specific completion standards, time to complete a flight lesson, and time to a successful evaluation flight. Percent transfer and transfer effectiveness ratios (TERs) were computed for each instrument task and for the time to complete a flight lesson. The data from the current study indicates that the FTD and the PCATD appear effective in teaching basic and advanced instrument tasks to private pilots but the limited number of subjects prevented this effectiveness from being convincingly demonstrated. As a result of prior training in an FTD and a PCATD time to a stage check or an instrument rating flight check flight was less when compared to an airplane control group.

Introduction

In an earlier study by Taylor et al., (1996), a commercially available Personal Computer Aviation Training Device (PCATD) was evaluated in a transfer of training experiment to determine its effectiveness for teaching instrument tasks. The data indicated that transfer savings for both the number of trials to reach a performance criterion for instrument tasks and time to complete a flight lesson were positive and substantial for new instrument tasks. A comparison of instrument rating course completion times resulted in a saving of about four hours in the airplane as a result of prior training in the PCATD. As a result of the Taylor et al. (1996) study, a Federal Aviation Administration (FAA) advisory circular published in 1997 permits 10 hours of instrument training to be completed in an approved PCATD.

To evaluate transfer of training effectiveness of a flight training device (FTD), the performance of subjects trained on instrument tasks in an FTD and later trained to criterion in an airplane must be compared to the performance of subjects trained to criterion only in the airplane. Roscoe (1971) demonstrated that the transfer effectiveness ratio (TER) accounts for the amount of prior training in ground trainers by specifying the trials/time saved in the airplane as a function of the prior trials/time in the ground training. Because diminishing transfer effectiveness ratios as the number of trials or hours in ground trainer increases, additional ground-based training will at some point cease to be cost effective. The law of diminishing returns adequately describes this relationship between extra training and resultant benefit. The purpose of the present study was to use an incremental transfer of training research design to measure the effectiveness of an FTD and a PCATD to determine the point at which additional training in a FTD or a PCATD is no longer effective.

Method

Participants

Participants were assigned to four FTD (Frasca) groups, one PCATD group, and a control (airplane) group. In the initial proposal a total of 180 pilots (30 in each of the 6 groups) were scheduled to participate in the study. Due to funding reductions in the second and third years, the number of pilots in the study was first reduced to a total of 120 pilots (20 subjects in each group) and due to the elimination of FY 2005 funding the eventual number of participants for each group who successfully completed the instrument program ranged between 15 and 20. The participants were University of Illinois, Institute of Aviation private pilot students, who were enrolled in the Institute's instrument flight program. This program consists of two semester courses: AVI 130, Basic Instruments and AVI140, Advanced Instruments. All students in the instrument program were involved in the study. A total of 106 students completed the study. Each semester the students were assigned equally to the six groups while maintaining a balanced number of subjects across all groups to account for students who did not complete the course prior to completion.

Equipment

Training in the FTD was conducted in four Frasca 141 FTDs with generic single-engine, fixed-gear, and fixed-pitch propeller performance models. The PCATD training was conducted using FAA approved PCATDs from Aviation Teachware Technologies (ELITE) v. 6.0.2, with flight controls by Precision Flight Controls. These PCATDs simulated the flight characteristics of the Piper Archer III aircraft. Airplane training was carried out in the Piper Archer III aircraft, which is a single-engine, fixed-pitch propeller, fixed undercarriage aircraft.

Procedure

The Frasca groups received 5, 10, 15, and 20 hours of prior instrument training in a FTD, respectively, and the PACTD group received 5 hours of prior training in the ELITE PCATD. .With the exception of the cross country training for Frasca groups 15 and 20 the prior training was distributed equally between AVI 130 and AVI 140. A Control group received all training in the airplane. Training on selected instrument tasks using the FTD and PCATD was administered to the four FTD groups and the PCATD group during four flight lessons for each semester. In addition, FTD training was given during certain x-country lessons in both AVI 130 and AVI 140 for the 15 and 20 hour FTD groups.

Prior to the start of each semester, all flight instructors were standardized on the use of the FTD and PCATD, changes in the training course outlines (TCOs), and experimental procedures. Flight instructors served as both instructors and data collectors. They rated student performances on designated flight tasks in the aircraft. For performance assessment in the aircraft, each instructor recorded if the student met the completion standards during the execution of the designated flight tasks. They also recorded the number of trials to criterion for specific tasks and flight time to complete a flight lesson (Phillips et al., 1995). Four check pilots, blind to the allocation of students to training conditions, were used to conduct the AVI 130 stage check and the AVI 140 instrument rating flight check.

Each flight instructor was instructed to schedule a stage check after Flight Lesson 40 in AVI 130, and an instrument rating flight check after Flight Lesson 55 in AVI 140 when the student was judged to be able to meet the proficiency standards for the stage check and the instrument proficiency check, respectively. These check flights permitted the assessment of the differential time to complete the flight course as a function of the amount of prior training in the FTD and the PCATD. Those students who failed the evaluation flight or failed to meet the proficiency standards by Flight Lesson 45 (stage check) and Flight Lesson 60 (instrument rating check flight) were provided additional flight time to reach proficiency. Dependent measures were trials in the airplane to proficiency, time to complete the flight lessons in the airplane, and total course completion time in the airplane for both courses.

Mean number of trials to reach criterion in the airplane for selected instrument tasks, and mean time to complete the flight lesson in the airplane were computed for all groups for both courses. Analyses of Variance (ANOVA) were performed to analyze the differences between the six groups. ANOVA were used to determine the significance of the trial variable and flight lesson completion time variable as a function of experimental treatment for both AVI 130 and AVI 140. Finally, ANOVA were used to determine the significance of the differences of the time to a successful check flight for the AVI 130 and AVI 140 courses as a function of the experimental treatment for the three groups (PCATD, FTD 5 and 10 groups) that received only prior training only on instrument tasks compared to the control group. To further identify the locus of any significant effects, post–hoc tests were used to make specific pairwise comparisons using Tukey's test of significance.

Results

A total of 124 subjects successfully completed the AVI 130 Basic Instruments course and took the final check ride. Table 1 shows the results of the check ride for the six groups. A total of 75 students passed the check ride on the first attempt and 49 students passed on the second attempt. Nine students were recommended for a remedial course, AVI 102. The total dual flight time to completion for the six groups is shown in Table 1 and in Figure 1. The average dual flight time to course completion for the airplane group was greater than the average time for each of the five experimental groups who had prior training in the PCATD or the FTD. The airplane group required 22.35 hours of dual to complete the course while the five experimental groups, after prior training in the PCATD or the FTD, required between 18.31and 20.87 hours of dual flight time in the airplane to complete the course.

For AVI 130, ANOVAs were computed to determine effect of the experimental treatment (assignment to groups) for mean trials to criterion in the airplane for selected instrument tasks for the four flight lessons for the three groups (PCATD, FTD 5 and 10 groups), that received prior training only on instrument tasks, and the control group. For Flight Lesson 37, there was a significant difference for both ILS and VOR ($F(3,81)=2.78$; $p < .05$ and $F(3,81)=5.12$; $p < .05$ respectively) and for Flight Lesson 38 there was a significant difference for VOR and DME ARC ($F(3,81)=2.84$; $p < .05$ and $F(3,81)=2.70$; $p < .05$ respectively). No other instrument tasks were significant. For Flight Lesson 37, pairwise comparisons using Tukey's test of significance indicated a significant difference between the airplane and the Frasca 5 and 10 groups ($p < .05$). ANOVA were computed to determine effect of the experimental treatment for mean time to complete the flight lesson for the four flight lessons for the PCATD, FTD 5 and 10 groups and the control group. A significant treatment effect was found for Flight Lessons 34/35, 36, and 37 (all $p < .05$). Pairwise comparisons indicated a significant difference between the airplane and all three groups for Flight Lesson 34/35 and between the Airplane and the Frasca 5 and 10 groups for Flight Lesson 37 (both $p < .05$). An ANOVA to determine effect of the experimental treatment for total course completion time in the airplane was computed. A significance difference was found (F (3,80)=3.67; p <.05. Pairwise comparisons using indicated a significant difference between the airplane and the Frasca 5 group ($p < .05$).

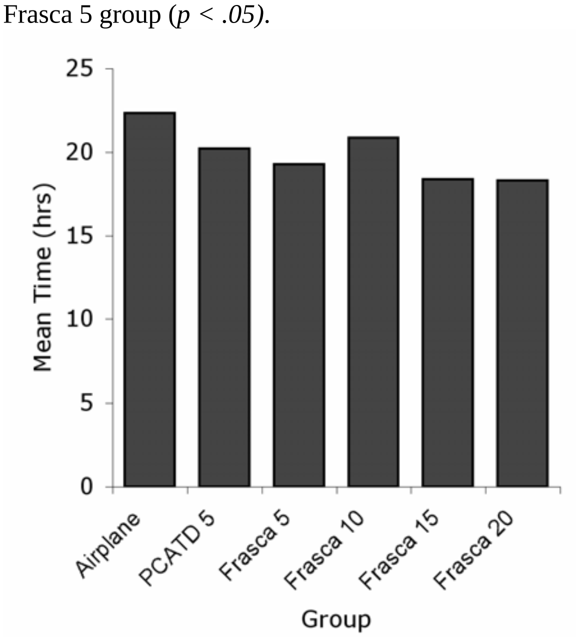


Figure 1. Total time to successful completion of flight lesson 45, showing incremental transfer effectiveness of the experimental groups.

A total of 106 subjects successfully completed the AVI 140, Advanced Instruments course and took the final check ride (the instrument rating flight check). Table 2 shows the results of the check ride. A total of 51 students passed the check ride on the first attempt and 46 students passed on the second attempt. The total dual flight time to completion for the six groups for the advance instrument course (AVI 140) is shown in Table 2 and in Figure 2. The average course completion time for the airplane group is greater for each of the five experimental groups who had prior training in the PCATD or the FTD. The airplane group required 26.38 hours of dual to complete the course while the total dual hours in the airplane to completion for the five experimental groups ranged from 25.78 to 20.79 hours after prior training in the PCATD or the FTD.

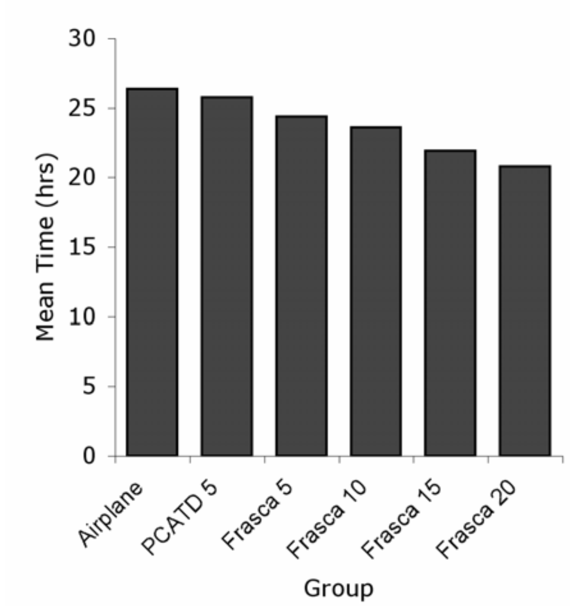


Figure 2. Total time to successful completion of flight lesson 60, showing incremental transfer effectiveness of the experimental groups.

For AVI 140, ANOVAs were computed to determine effect of the experimental treatment (assignment to groups) for mean trials to criterion in the airplane for selected instrument tasks for the four flight lessons for the three groups (PCATD, FTD 5 and 10 groups), that received prior training only on instrument tasks, and the control group. For Flight Lesson 48, there was a significant difference for ILS approach (F (3.77)=2.90; $p < .05$). Pairwise comparisons indicated a significant difference between the PCATD 5 and the Frasca 5 group ($p < .05$). For Flight Lesson 50, there was a significant difference for NDB approach (F (3,77)=3.90; $p < .05$). Pairwise comparisons indicated a significant difference between the Airplane and the PCATD 5 and the Frasca 5 groups ($p < .05$). For Flight Lesson 52, there was a significant difference for NDB Hold and GPS approach (F (3,76)=3.34; $p < .05$ and F (3,75)=3.14; $p < .05$ respectively). Pairwise comparisons indicated a significant difference between the PCATD 5 and the Frasca 5 groups for NDB Hold ($p < .05$). ANOVAa were computed to determine effect of the experimental treatment for mean time to complete the flight lesson for the four scored flight lessons for each of the three groups (PCATD, FTD 5 and 10 groups) that received only prior training on instrument tasks and the Control group. A significant treatment effect was found for Flight Lesson 52 (F (3,76)=5.79; $p < .05$). Pairwise comparisons indicated a significant difference between the PCATD 5 and the Frasca 5 and 10 groups ($p < .05$). An ANOVA was computed to determine effect of the experimental treatment for total course completion time in the airplane for AVI 140. A significance difference was found (F (3,65)=2.77; $p < .05$). Pairwise comparisons indicated no significant difference between any groups.

The effect of allocating 5 and 10 hours in the Frasca for cross-country flight was evaluated. For AVI 140, the airplane group required 26.38 hours of dual to completion while the Frasca 10,15 and 20 groups required 23.60, 21.93 and 20.79 hours respectively. This represents a savings of 2.78 hours, 4.45 hours and 5.59 hours respectively. Since the Frasca 15 and 20 groups received the same treatment as the Frasca 10 group regarding training only on instrument tasks and an additional 5 and 10 hours respectively for cross country training, the computed savings for the 5 and 10 hours cross country time was 1.67 and 2.81 hours respectively.

Discussion

The data from the current study indicates that the FTD and the PCATD appear effective in teaching basic and advanced instrument tasks to private pilots but the limited number of subjects prevented this effectiveness from being convincingly demonstrated. With the limited number of subjects and the current variability among subjects, the power of the ANOVA is low. The current data fail to replicate the findings of Taylor et al. (1996, 1999) that PCATDs are useful to teach instrument tasks to private pilots. As a result of prior training in an FTD and a PCATD, time to the stage check in AVI 130 and to the instrument rating flight check was less for three groups (PCATD, FTD 5 and 10 groups) that received prior training only on instrument tasks as compared to the control group. For AVI 130, pairwise comparisons indicated a significant difference between the airplane and the Frasca 5 group and for AVI 140, pairwise comparisons indicated no significant difference between any groups. One purpose for conducting an incremental transfer of training study is to determine at what point additional training in the FTD and the PCATD is no longer effective. The data collect does not permit this to be determined convincingly. A study by Taylor et al., (2002) clearly indicated that the use of 5 hours of PCATD time was cost-effective based on the allocation of PCATD time for these tasks for the PCATD 5 group. The current study shows that the PCATD is only effective for the NDB task. We attribute the difference between the two studies to be the result of the lack of power in the current study.

Time to complete the flight lesson was significant for three flight lessons out of four for AVI 130 when comparing the PCATD, FRASCA 5 and 10 groups with the Control group, but for only one flight lesson out of four for AVI 140. Taylor, et al (2002), which tested the incremental effectiveness of the PCATD, found two of four flight lessons significant for AVI 130 and one for AVI 140.

We do not believe that data generated in the current study provides convincing evidence for flight schools to use in determining how to best implement PCATDs or FTDs in their training programs. There is the possibility that FTDs can be used effectively for teaching cross-country procedures in addition to using them to teach instrument tasks, but the current study has failed to demonstrate significant savings through their use.

Acknowledgments

This work was supported under Federal Aviation Administration (FAA) Cooperative Agreement Number 02-G-033 and sponsored by the FAA Headquarters, Flight Standard Service, General Aviation & Commercial Division. Dennis Beringer served as the COTR. Views expressed herein do not necessarily represent official FAA positions. We express our appreciation to Ms. Sybil Phillips for invaluable assistance with flight operations and with student management. Mr. Bill Jones, Mr. David Boyd, and Don Talleur served as check pilots. We also thank the Institute of Aviation flight instructors and students for their participation in the study.

References

Phillips S. I., Taylor, H. L., Lintern, G., Hulin, C. L., Emanuel, T., & Talleur, D. (1995). Developing performance measures for evaluating personal computer-based aviation training devices within a FAR Part 141 pilot training school. *Proceedings of Aviation Psychology 8th International Symposium*. Columbus, OH: Ohio State U.

Roscoe, S. N. (1971). Incremental transfer effectiveness. *Human Factors, 13*, 561-567.

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D., Emanuel, T., & Phillips, S. (1996). *Transfer of training effectiveness of personal computer-based aviation training devices* (ARL-96-3/FAA-96-2). Savoy, IL: Aviation Research Laboratory.

Taylor, H. L., Lintern, G, Hulin, C. L., Talleur, D. A., Emanuel, T., & Phillips, S. (1999). Transfer of training effectiveness of a personal computer aviation training device. *International Journal of Aviation Psychology, 9*, 319-335.

Taylor, H. L., Talleur, D. A., Emanuel, T. W., Jr., Rantanen, E. M., Bradshaw, G. L., & Phillips, S. I. (2002). Incremental training effectiveness of personal computers used for instrument training (Final Technical Report ARL-02-5/NASA-02-3). Savoy, IL: University of Illinois, ARL.

Table 1.
Flight Lesson 45 Statistics (Fall, 2002, Spring, Summer, Fall 2003 and Spring 2004)

	Airplane Only	PCATD 5.00	Frasca 5.00	Frasca 10.00	Frasca 15.00	Frasca 20.00
Number of Students	22	20	22	20	21	19
% First Flight Pass Rate	59.00 (N=13)	65.00 (N=13)	45.45 (N=10)	75.00 (N=15)	76.19 (N=16)	42.11 (N=8)
% Second Flight Pass Rate	100.00 (N=9)	100.00 (N=7)	100.00 (N=12)	100.00 (N=5)	80.00 (N=5)	100.00 (N=11)
Students Recommended 102	0	0	1	1	4	3
Total Dual to Completion	22.35 (N=22)	20.20 (N=20)	19.27 (N=22)	20.87 (N=20)	18.36 (N=21)	18.31 (N=19)
Variance Tot. Dual to Completion	9.39	6.40	10.03	14.17	9.87	9.48

Note: This lesson is the final check ride for AVI 130.

Table 2.
Flight Lesson 60 Statistics (Spring, Summer, Fall, 2003, Spring, Summer, Fall 2004)

	Airplane Only	PCATD 5.00	Frasca 5.00	Frasca 10.00	Frasca 15.00	Frasca 20.00
Number of Students	18	18	20	16	15	19
% First Flight Pass Rate	44.44 (N=8)	55.56 (N=10)	45.00 (N=9)	43.75 (N=7)	40.00 (N=6)	57.89 (N=11)
% Second Flight Pass Rate	100.00 (N=10)	75.00 (N=6)	88.89 (N=8)	88.89 (N=8)	100.00 (N=9)	62.50 (N=5)
Students Recommended 102	2	3	4	3	5	2
Total Dual to Completion	26.38 (N=18)	25.78 (N=17)	24.40 (N=18)	23.60 (N=16)	21.93 (N=15)	20.79 (N=18)
Variance Tot, Dual to Completion	16.55	6.03	7.92	8.80	10.20	17.89

Unmanned Aircraft Pilot Medical and Certification Requirements

Kevin W. Williams, Ph.D.
FAA Civil Aerospace Medical Institute, Oklahoma City, OK

ABSTRACT

A research effort was undertaken to establish unmanned-aircraft pilot medical and certification requirements. The effort consisted of a review of relevant literature, a summary of potential unmanned aircraft applications, a review of proposed applications by members of RTCA SC-203, the convening of a panel of subject matter experts, and interactions with groups engaged in the process of establishing unmanned aircraft pilot guidelines. The results of this effort were a recommendation and justification for use of the Class III medical certification and recommendations regarding the training and testing of unmanned - aircraft pilots.

INTRODUCTION

The rapidly expanding commercial Unmanned Aircraft (UA) industry presents a challenge to regulators whose task it is to ensure the safety of the flying public as well as others who might be injured as a result of an aircraft accident. The military has used unmanned aircraft for several decades with various levels of success. Within the last few years, commercial UA operations have increased dramatically. Most of these operations have concentrated on surveillance and advertisement, but several companies have expressed an interest in using unmanned aircraft for a variety of other commercial endeavors.

Although the term "unmanned aircraft" suggests the absence of human interaction, the human operator/pilot is still a critical element in the success of any unmanned aircraft operation. For many UA systems, a contributing factor to a substantial proportion of accidents is human error (Williams, 2004). The FAA needs guidance to assist in the decision of who will pilot UA and what type of training will be required. Research may be required: to investigate the effects on pilot performance of different types of console display interfaces; to determine how UA flight mission profiles affect pilot workload, vigilance, fatigue, and performance; to determine whether prior flight experience is important to operate a UA; to determine whether new opportunities present themselves in terms of the inclusion of persons with handicaps that were previously excluded from piloting aircraft but would not have difficulty with UA; and to investigate medical and physiological standards required to operate a UA.

To assist in developing guidance, an effort was begun to study UA pilot medical and certification qualifications. The approach consisted of several steps. First, a literature review of existing research on UA pilot requirements was conducted. Second, analyses of current and potential UA commercial applications and of current and potential UA airspace usage were completed. The third step in the process was the assembling of a team of subject matter experts that reviewed currently proposed UA pilot medical and certification requirements and made recommendations regarding how those requirements should be changed or expanded. This information, along with the other efforts, was used to develop preliminary task analyses of the unmanned aircraft piloting task. This paper is a summary of this effort.

UA Pilot Requirements Literature Review

The first task was to conduct a review of literature related to the development of UA pilot requirements. The literature fell into just a few basic categories. Many of the papers were recommendations regarding the development of requirements (e.g., DeGarmo, 2004; Dolgin, Kay, Wasel, Langelier, & Hoffman, 2001; Reising, 2003). The paper by Weeks (2000) listed current crew requirements for several different military systems. Finally, some of the papers were a reporting of actual empirical research addressing some aspect of pilot requirements (Barnes & Matz, 1998; Fogel, Gill, Mout, Hulett, & Englund, 1973; Schreiber, Lyon, Martin, & Confer, 2002).

The research by Fogel et al. (1973) was especially interesting because it was one of the earliest attempts to address the issue of UA pilot requirements. In the study, three groups of pilots were recruited to fly a simulation of a Strike remotely piloted vehicle. The first group consisted of Navy attack pilots with extensive combat aircraft experience. The second group consisted of radio-control aircraft hobbyists. The third group was composed of non-pilots with no radio-control aircraft experience. The results showed that, even though the Navy pilots were better than either of the other two groups, the other groups showed significant improvement in flight control over the course of the sessions, leading the authors to state, “It is hypothesized that a broader segment of relatively untrained personnel could be brought up to the required level of skill with short time simulation/training provided they meet some minimum selection criteria” (Fogel, et al., 1973, p. 75). It should be noted that the control interface consisted of a joystick for controlling the aircraft (but no rudder pedals), with very little in the way of automation for simplifying the control task. However, the researchers did compare two types of flight control systems, with the joystick either directly controlling (simulated) aircraft surfaces or a more sophisticated control system where the joystick commanded the aircraft performance (bank and pitch) directly. The authors concluded that the performance control joystick was superior for aircraft control, regardless of the level of pilot experience.

The research by Schreiber et al. (2002) looked at the impact of prior flight experience on learning to fly the Predator UAS. Seven groups of participants were used in the study, ranging from no flight experience to prior Predator flight experience. Results showed that the group with no flying experience performed significantly worse than the other groups, while the group with previous Predator experience performed significantly better. This finding was expected. However, an unexpected finding from the study was that participants with various levels and types of non-Predator flight experience all performed relatively the same with the Predator system. The authors concluded that any type of flight experience with an aircraft with similar handling characteristics to the Predator was beneficial for flight training on the Predator system. The authors pointed out, though, that the study looked only at stick and rudder skills, and not at more general types of flight skills such as communication and airspace management. In addition, the study did not address whether other types of training, such as simulator training, would also be useful for the transfer of Predator flight skills.

While it might be possible to establish whether a certain type of training or experience is more effectively transferred to a particular UA system, such as the Predator, these studies have not answered the question of whether manned aircraft time is required to be a successful pilot of an unmanned aircraft. We know that certain systems, such as the U.S. Army Hunter and Shadow systems, are successfully flown by pilots with no manned-aircraft experience. However, once these systems begin flying in populated airspace, there is a question of whether a lack of manned-aircraft experience within the airspace might degrade the effectiveness of the pilot and the safety of the flight. Research is needed to address this issue.

UA Applications and Airspace Usage

For a summary of UA applications and airspace usage issues, please reference the technical report (Williams, in review).

Summary of Meeting on UA Pilot Medical and Certification Requirements

On July 26th, 2005, a meeting was held at the FAA Civil Aerospace Medical Institute (CAMI) in Oklahoma City, OK. The purpose of the meeting was to assemble a diverse group of subject matter experts, from industry, academia, the FAA, and the military, to discuss Unmanned Aircraft (UA) pilot medical and certification requirements.

Attendees included representatives of several groups currently working on the development of standards and guidelines for UA. There were representatives from NASA Access 5, ASTM F38, RTCA SC-203, and SAE-G10 at the meeting. In addition, Dr. Warren Silberman represented the FAA Airmen Medical Certification Division and the Office of Aviation Medicine in regard to the medical certification requirements discussion.

Because the meeting was for only one day, an attempt was made to focus the discussion as much as possible by providing a draft standard that was developed by the Flight Standards Division (AFS-400). In particular, two paragraphs from the draft UA standards were reviewed and discussed extensively during the meeting. These two paragraphs are shown below.

6.14 **Pilot/Observer Medical Standards**. Pilots and observers must have in their possession a current third class (or higher) airman medical certificate that has been issued under 14CFR67. 14CFR91.17 regulations on alcohol and drugs apply to both UA pilots and observers.

6.15 **Pilot Qualifications**. The intent of this paragraph is to ensure that UA pilots interacting with ATC have sufficient expertise to perform that task readily.

6.15.1 Pilots must have an understanding of Federal Aviation Regulations applicable to the airspace where the UA will operate.

6.15.2 If the UA is operating on an instrument flight plan, the UA pilot must have an instrument rating.

6.15.3 Pilots flying UA on other than instrument flight plans must pass the required knowledge test for a private pilot certificate as stated in 14CFR61.105 (or military equivalent) for all operations beyond visual line-of-sight and for all operations conducted for compensation or hire regardless of visual proximity.

6.15.4 Pilots requiring instrument ratings will be certificated pilots of manned aircraft.

6.15.5 Equivalent military certificates and training are acceptable in all cases.

In the end, it was decided that not enough was known about these aircraft to make an accurate assessment of all of the risks involved. Because of this, the decision was reached by the group that the original suggestion of a class III medical certification was good, with use of the existing medical waiver process for handling exceptions (e.g., paraplegics). This decision is also supported by the factors identified above that mitigate the severity of pilot incapacitation. However, there was some additional discussion that some applications might require a class II or I medical certification because of the increased risks involved. Imposing different certification requirements, though, would require a clearer specification of pilot certification levels and UA classes. The class III medical certification statement was believed to apply to many, if not all, existing commercial and public UA endeavors (public endeavors would include border patrol applications). The question thus arose as to what types of pilot certification would require a stricter medical certification. Since the document was viewed as certainly undergoing revisions in the future, no wording changes were suggested at this time for paragraph 6.14.

A complete summary of the meeting can be found in the technical report (Williams, in review).

Identification of Knowledge, Skills and Abilities

One final effort undertaken in the research this year was the development of a set of knowledge, skills, and abilities required by the UA pilot. Several groups are working on the development of pilot KSAs, including NASA Access 5 and SAE-G10. The KSAs that have been developed are very similar across the groups because they rely heavily on manned aircraft tasks.

There are, however, three areas that have been identified that distinguish manned from unmanned aircraft. These areas will be important during the development of training and test standards for these systems. The areas are 1) activities and information related to the data link, 2) activities and information related to the task of detecting, sensing, and avoiding aircraft, and 3) activities and information related to the handoff of control during the flight.

Data link issues cut across the entire flight, from pre-flight planning until recovery of the aircraft. It is important that the pilot have an understanding of the conditions that affect the data link during the flight, and be prepared to take appropriate action if the data link is lost. During pre-flight, the pilot should be aware of the weather conditions that will occur during the flight and understand how those conditions will affect the data link. The pilot must also know which portions of the flight might be susceptible to interference or blockage of the data link due to natural barrier or broadcasting. There should also be contingency plans during each leg of the flight in case of a loss of data link. During the flight, there should be procedures for attempting to re-establish the data link if it is lost, and for notifying others, such as air traffic control, if the data link cannot be re-established.

There should be established procedures for detecting, sensing, and avoiding other aircraft during the flight. These procedures might begin before the flight, with the notification of other traffic that an unmanned aircraft will be flying in the airspace. The limitations of whatever method is in place for detecting other aircraft should be well understood. Also, the procedures for avoiding aircraft should be understood and practiced before they have to be used.

The handoff of control during a flight will be a common occurrence for a great many UA systems. Control handoff can occur in a variety of ways. Each method introduces the possibility of human error and has been the cause of a variety of UA accidents (Williams, 2004).

SUMMARY AND CONCLUSIONS

There were two goals for the research that was conducted. The first was a specification of the medical requirements for UA pilots. The second was a specification of the certification requirements for UA pilots.

The establishment of medical requirements for UA pilots was based on an analysis of the method for establishing the medical requirements of other occupations, including manned aircraft pilot. Rather than suggesting the creation of a new medical certification for UA pilots, it was decided to use an existing pilot medical certification. There were several reasons supporting this decision, including the bureaucratic difficulty in establishing a new certification level and the problems associated with training medical examiners who would be asked to assess whether pilots successfully met the new requirements.

Given that an existing medical certification was to be used, the question of which level of certification should be required was then based on the perceived level of risk imposed by the potential incapacitation of the UA pilot. The third class medical certification was judged to be the most acceptable based on the idea that there were several factors that mitigated the risk of pilot incapacitation relative to manned aircraft. First, factors related to changes in air pressure could be ignored, assuming that control stations for non-military operations would always be on the ground. Second, many of the current UA systems have procedures established for lost data link. Lost data link, where the pilot cannot transmit commands to the aircraft, is functionally equivalent to pilot incapacitation. Third, the level of automation of a system determines the criticality of pilot incapacitation, since some highly automated systems (e.g., Global Hawk) will continue normal flight whether a pilot is present or not.

The specification of certification requirements for UA pilots should be based on a task analysis of the UA piloting task and a specification of the knowledge, skills, and abilities needed for the task. While several groups have been working on completing such a task analysis, the work is still ongoing. Therefore, it is not possible at this time to reach definitive conclusions regarding certification requirements for UA pilots.

The available research on pilot qualifications shows that, while manned-aircraft experience is beneficial for piloting some UA systems (Schreiber et al., 2002), basic stick-and-rudder skills can also be mastered by those without flight experience (Fogel et al., 1973). This, of course, makes sense since even pilots with manned-aircraft experience had no flight experience at some point in their career. The question in regard to whether or not manned-aircraft flight experience should be a prerequisite for UA pilots centers on whether there is any learning that occurs during manned-aircraft flight training that would not be adequately addressed during training with an unmanned aircraft. One possibility is the idea of "shared fate". The fact that the pilot does not share the fate of the aircraft might lead to differences in decision-making during a flight (McCarley & Wickens, 2005). Another possibility, though one that has not been addressed experimentally, is that a full understanding of the three-dimensional aspect of the aircraft in the airspace cannot occur without experience in the airspace. Research is required to address this issue.

An analysis of the types of applications expected for UA indicated that airspace usage might be neatly divided between applications that use only Class G airspace and those that use other classes. Those that use only Class G airspace, with the exception of flights within restricted areas such as military areas of operation, were limited to line-of-sight from the pilot. Those that utilized other classes of airspace were always beyond-line-of-sight. This distinction (line-of-sight vs. beyond-line-of-sight) might be a useful way to classify types of unmanned aircraft for purposes of airworthiness ratings as well as pilot ratings.

Finally, while both training and test standards should be structured similarly to manned aircraft training and testing, they should include areas that are unique to the piloting of unmanned aircraft. Three areas that were identified as unique were data link issues, detect, sense, and avoid issues, and control handoff issues. The development of training and testing standards will require that these issues be addressed completely.

REFERENCES

Barnes, M.J. & Matz, M.F. (1998). Crew simulations for unmanned aerial vehicle (UAV) applications: Sustained effects, shift factors, interface issues, and crew size. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 143-7.

DeGarmo, M. (2004). *Issues concerning integration of unmanned aerial vehicles in civil airspace.* Mitre Corporation Report # MP04W0000323.

Dolgin, D., Kay, G., Wasel, B., Langelier, M., & Hoffman, C. (2001). Identification of the cognitive, psychomotor, and psychosocial skill demands of uninhabited combat aerial vehicle (UCAV) operators. Downloaded from URL http://forum.nomi.med.navy.mil/articles/safeucav/ on 3/10/2005.

Fogel, L.J., Gill, R.S., Mout, M.L., Hulett, D.G., & Englund, C.E. (1973). *Principles of display and control design for remotely piloted vehicles.* Decision Science, Inc. second semi-annual technical report on Contract #N00014-72-C-0196, Project # NR 196-119/11-29-71 455.

McCarley, J.S. & Wickens, C.D. (2005). *Human factors implications of UAVs in the national airspace*. University of Illinois Institute of Aviation Technical Report (AHFD-05-5/FAA-05-1). Savoy, IL: Aviation Human Factors Division.

Reising, J.M. (2003). The role of operators in uninhabited military vehicles: A NATO perspective. *Proceedings of the 12th International Symposium on Aviation Psychology*, Dayton Ohio, April 14-17, 988-92.

Schreiber, B. T., Lyon, D. R., Martin, E. L., & Confer, H. A. (2002). *Impact of prior flight experience on learning Predator UAV operator skills* (AFRL-HE-AZ-TR-2002-0026). Mesa, AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Weeks, J.L. (2000). *Unmanned aerial vehicle operator qualifications (AFRL-HE-AZ-TR-2000-0002)*. Mesa, AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Williams, K.W. (2004). *A summary of unmanned aircraft accident/incident data: Human factors implications.* U.S. Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine, Washington, DC. Technical Report Publication No. DOT/FAA/AM-04/24.

Williams, K.W. (in review). *Unmanned aircraft pilot medical and certification requirements*. U.S. Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine, Washington, DC. Technical Report.

USE OF WEATHER INFORMATION BY GENERAL AVIATION PILOTS: PROVIDERS AND PRODUCTS

William R. Knecht
Civil Aerospace Medical Institute
Oklahoma City, OK

Data obtained from 211 general aviation (GA) pilots were examined to determine usage patterns for weather information. Weather products, providers, and en-route information sources were ranked according to relative use and rated by perceived information value, frequency of use, and time invested per usage. The measures were highly correlated. Additionally, voice tapes of 306 calls to Automated Flight Service Stations were analyzed. Conclusion #1: A small fraction of pilots show sparse use patterns. These may be at risk for flying with inadequate preparation. Conclusion #2: There seems to be a strong tendency for many pilots to prefer relatively simple forms of information (e.g. METARS). This may present a problem, given the often-complex nature of weather.

INTRODUCTION

Purpose

This report is a summary of onsite surveys completed by GA pilots concerning their use of weather information products and providers. The intent was to establish actual usage of products and services as compared with the recommended strategy for using weather data.

GA weather products and providers

A weather *product* is a relatively small package of related information constituting a stand-alone report (e.g. METAR, TAF). Weather *providers* are organizations dedicated to bundling weather products, NOTAM, TFR, and flight planning information into convenient, user-friendly form. The Flight Service Station is a good example of a weather provider. Providers try to give us a strategic sense of the weather, to complement the tactical sense given by the separate weather products themselves. There are literally scores of weather providers, most of them commercial, for-profit. Many high-end providers offer features rivaling those available to airline dispatchers.

METHOD

Design and participants

During July and August 2005, the FAA Civil Aerospace Medical Institute surveyed over 230 GA pilots at locations across 5 states (CA, OK, ND, IL, FL). Pilots' median age was 23 years and median flight experience was 245 hours. Women made up 14% of the sample. All were volunteers paid for their services as SMEs.

Procedure

Pilots were asked to a) rate weather products and providers on the basis of how much they typically used them, b) assign each a value based on its information content, c) estimate the percentage of times each was used on a "standard flight" and, d) estimate the number of minutes each was used on this standard flight. This "standard flight" was defined as a 4-hour flight through "weather serious enough to challenge *your* skill level and the aircraft's capabilities."

RESULTS

Weather providers

For each weather provider, Table 1 shows four ratings supplied by pilots plus one rating arithmetically derived from the last two ratings. *Rank* is a rank-ordering of how much pilots felt they used a given weather provider. *Value* is a similar measure of how valuable each pilot felt that provider's information was. *Percent of Flights* refers to the percentage of flights on which pilots used each provider. *Minutes Spent* refers to the amount of time per flight a given provider was used if and when it was accessed. The final column, *Average Minutes Spent* per flight, is the result of multiplying *Percent of Flights* times *Minutes Spent*. As such, *Average Minutes Spent* is a statement about how much time was spent on a given provider on the "average" flight (even though sometimes it may have been used and sometimes not). *Average Minutes Spent* also allows us to estimate an average *Total Minutes Spent Per Flight* across all providers (19.8 in this case).

Note that *Rank* does not have to equal *Value*. For instance, we might highly *value* a Rolls-Royce automobile, yet *rank* it low in terms of use, since we cannot afford to actually own one.

Ranks and values were all normalized to a scale of zero to one to allow for easier comparison of the data. Directionality is such that "0" represents *least* valuable and "1" represents *most* valuable.

Notice that the FSS standard briefing was both ranked and valued highest (1.0), and said to be used on the highest percent of flights (61%). This was closely followed by the public NWS/NOAA Web sites, which actually experienced higher minutes-spent-when-used and overall average minutes used.. Internet DUATS also received high ratings across the board.

Table 1. Normalized ranks, values, frequency of use and time using various weather information sources.

Provider	Format	Rank 0-1	Value 0-1	% of flights	Min. spent	Ave. min spent
Commercial vendor	Internet	0.4	0.5	28.7	5.0	1.4
Public NWS or NOAA site	Internet	0.7	0.8	49.8	**13.9**	**6.9**
DUATS	Internet	0.7	0.7	34.0	8.9	3.0
DUATS	at airport	0.1	0.1	11.3	2.1	0.2
FSS (automated TIBS)	telephone	0.1	0.1	8.9	1.5	0.1
FSS (standard)	telephone	**1.0**	**1.0**	**61.5**	9.1	5.6
FSS (abbreviated)	telephone	0.1	0.2	9.2	1.8	0.2
FSS (outlook)	telephone	0.2	0.3	14.4	2.4	0.3
The Weather Channel	Internet,TV	0.4	0.5	27.9	7.0	2.0

Finally, a surprising number of pilots reported using The Weather Channel, even though it is not an FAA-approved source. This was perhaps due to the sheer ease of turning on the television and watching TWC during morning coffee. Also, Internet TWC has a convenient feature allowing the user to type in a zip code and receive easy-to-understand forecasts based on current location. In other words, TWC seems to give pilots something they want—a simple report, local and fast. The other sources are far more comprehensive, but that breadth comes at the expense of extra time and effort.

Weather products

As stated, *Rank* and *Value* were normalized here, so direct comparisons can be made between all four categories of responses. Table 2 shows that the six most highly ranked and valued weather products were TAF and METAR (tied for first place), followed by AIRMET/SIGMET, radar charts, FAs, and ATIS, all more or less tied for second. The total estimated average number of minutes spent was 16.6, reasonably consistent with the 19.8 estimated for time spent using providers (the importance of this will be discussed in greater detail later).

It is quite interesting that the "old standbys," METAR and TAF, rated so highly. Again, this may parallel TWC's popularity in some human tendency to want brevity and simplicity. As human factors researchers, we would all be well-advised to remember this psychological principle.

En-route sources

Table 3 shows results for the en-route information sources. Two relatively simple sources - ATIS and AWOS - tied roughly for first place. ASOS and Flight Watch tied for second. It was not obvious why Flight Watch did not receive higher ratings. Perhaps it was merely because this was a "one-tank" flight. Had we specified, say, a longer flight requiring refueling, perhaps we would have seen a shift in the numbers. The relatively low ranking of weather-related avionics may stem from several sources. It is possible that access to in-flight sources is more likely with the older more established pilots who own aircraft or have invested in portable devices with subscriptions to data-providing up-link services. The average age of the present sample was comparatively young and they might be less likely to have access to these sources.

Table 2. Normalized ranks, values, frequency of use and time using various weather products.

Product	Format	Rank 0-1	Value 0-1	% of flights	Min. spent	Ave. min spent
AC (Severe Wx Outlook Narrative)	text	0.1	0.1	4.7	0.4	0.0
AIRMET / SIGMET	text	0.5	0.7	47.6	3.7	1.8
ASOS (Automated Surface Observing System)	radio	0.2	0.2	13.0	0.8	0.1
ATIS (Automated Terminal Information System)	radio	0.4	0.5	41.4	2.0	0.8
AWOS (Automated Weather Observing System)	radio	0.3	0.4	25.0	1.8	0.5
charts, Air- or Surface-analysis	graphic	0.1	0.2	12.8	1.0	0.1
charts, Convective outlook	graphic	0.1	0.1	10.1	1.1	0.1
charts, Prog.	graphic	0.2	0.3	17.8	1.7	0.3
charts, Radar (NEXRAD)	graphic	0.5	0.6	44.2	3.6	1.6
charts, Radar summary	graphic	0.3	0.4	23.7	1.7	0.4
charts, Weather depiction	graphic	0.2	0.3	15.1	1.8	0.3
FA (Aviation area 18-hr forecast)	text	0.5	0.5	36.1	3.2	1.2
FD (Winds and temps aloft)	text	0.3	0.4	30.0	2.2	0.7
FD	graphic	0.0	0.1	5.3	0.4	0.0
GPS (Global Positioning Satellite)	T or G	0.1	0.1	5.1	0.5	0.0
LLWAS (Low Level Wind shear Alerting System)	radio	0.0	0.0	0.9	0.1	0.0
METAR	text	**1.0**	**1.0**	**77.3**	**4.5**	**3.4**
PIREP	text	0.3	0.6	36.4	2.2	0.8
Satellite (images of cloud cover)	graphic	0.2	0.3	20.9	1.8	0.4
SD (hourly weather reports)	text	0.0	0.0	3.9	0.4	0.0
TAF	text	**1.0**	**1.0**	**76.5**	**5.3**	**4.0**
TWEB (Transcribed Weather Broadcast)	radio	0.1	0.1	9.0	0.9	0.1
WW, AWW (weather watch bulletins)	text	0.0	0.0	0.1	0.1	0.0
Other sources		0.0	0.0	0.1	0.0	0.0
		Total minutes spent per flight				**16.6**

Table 3. Normalized ranks, values, frequency of use and time using various enroute weather sources.					
	Rank	Value	% of	Min.	Ave. min
En-route source	0-1	0-1	flights	spent	spent
avionics	0.1	0.0	8.3	1.2	0.1
ASOS	0.3	0.4	23.6	1.6	0.4
ATIS	**1.0**	**1.0**	75.6	4.6	3.5
AWOS	0.6	0.7	48.7	4.1	2.0
EFAS (FSS Flight Watch)	0.4	0.6	29.1	4.1	1.2
HIWAS (Hazardous Inflight Weather Advisory System)	0.2	0.3	14.0	1.4	0.2
TWEB	0.0	0.0	2.6	0.4	0.0
Other sources	0.0	0.0	4.0	0.3	0.0

Analysis of Voice Recordings from Automated Flight Service Stations (AFSS). One additional focused examination was conducted for the pilots' first-choice information source, the Automated Flight Service Stations (AFSS). The interest was in determining what types of information AFSS specialists provide, what pilots request, and how they might use that information. To answer these questions, three AFSS facilities provided 24 hours of continuous recordings of actual recent conversations that occurred between pilots and specialists staffing the preflight desk. The recordings represented 306 calls made on good (90), typical (80), and poor (136) weather days occurring in the Northwest Mountain Region (95), Southwest Region (105), and New England Region (106). Data extracted from the tapes included whether the pilot requested (259) or declined (47) a preflight briefing and the types of weather information pilots requested or that were provided by specialists. The pilots who called fell into 3 basic groups: (1) local fliers; training schools, students, and aircraft buffs who stay within 30-50 miles of the departure point and return to that airport, (2) fixed base operators (FBO) who rent aircraft and transport passengers for hire, advanced training, and short distance carriers (with stored or pre-filed flight plans), and pilots of larger aircraft, and (3) business, military (training and operations), corporate, and long-distance lifeguard pilots.

Generally, pilots requested standard weather briefings more often (VFR 43%, IFR 37%) than either abbreviated (VFR 38%, IFR 27%) or outlook (VFR 8% IFR 6%) briefings. Regardless of weather conditions, AFSS relayed the following weather items in 85% of the pilot-requested preflight weather briefings: Weather synopsis, sky conditions (clouds), visibility, weather conditions at the departure, en route, and destination point. Also included to a lesser degree were adverse conditions, altimeter setting, cloud tops, dew point, icing conditions, surface winds, winds aloft, temperature, thunderstorm activity, precipitation, precipitation intensity, visibility obscuration, other weather, PIREPs, AIRMETs/SIGMETs, MOAs, MTRs, NOTAMs, and TFRs.

During typical weather conditions, pilots who did not request a preflight briefing still asked the specialist about the weather conditions at their departure point (25%), en route (25%) and at the destination point (25%). On marginal VFR days they also asked about any TFRs, NOTAMs, AIRMET/SIGMET and PIREP as well as thunderstorm activity, winds aloft, cloud tops, and ATC delays or flow control advisories. Whether by asking for additional information or receiving weather information from specialists, 32 pilots decided that it was best to change their flight plans. Some delayed (47%), postponed or cancelled (16%) their flights while others looked for alternate routes and destination points (16%). It was not immediately evident why pilots declined the weather briefing in 15% of the calls, but it could be speculated that currently available weather-information sources such as internet aviation weather services and DUATS allowed these pilots to be comfortable with the information from these sources in lieu of a preflight briefing by a specialist.

Reliability and Consistency of the data

One measure of reliability may be obtained by comparing the time pilots said they spent on weather products versus on providers (16.6 vs. 19.8 minutes). Since providers consist of products plus other services, the number associated with providers should be close to, but slightly greater than, that for products. As expected, that is the case.

Intercategory correlations can also be used to infer some measure of reliability. If data categories are designed so that multiple questions are asked about similar things then, if respondents give logically consistent answers across categories, it can be assumed that most were answering items thoughtfully rather than randomly. *Rank*, *Value*, *Percent Use*, and *Minutes Used* all logically measured related aspects of value to pilots. Therefore, they should all correlate as long as participants did not answer randomly.

In Table 4 we do see very high groupwise intercategory correlations, ranging from 89-99%. From this we can infer a number of things. First, we have at least some indication that our data are reliable. If we did the same study again with the same pilots, we ought to get similar results.

Table 4. Provider, product, and enroute source intercorrelations.

	Provider intercorrelations					Product intercorrelations					En-route source intercorrelations			
	Rank	Value	%	Min.		Rank	Value	%	Min.		Rank	Value	%	Min.
Rank	1					1					1			
Value	0.993	1				0.975	1				0.979	1		
%	0.988	0.987	1			0.987	0.993	1			0.994	0.961	1	
Min.	0.896	0.910	0.902	1		0.954	0.972	0.966	1		0.927	0.960	0.898	1

Second, the high intercorrelations imply: a) Pilots do generally seem to use the information they value most (unlike our example with the Rolls-Royce), and; b) In future studies it is probably unnecessary to use all four measures. *Percent Use* and *Minutes Used* are probably sufficient, both to check reliability and to estimate the total minutes each pilot spends on weather briefings.

Table 5. Ranks, values, frequency of use, time used, and estimated total times involved in weather briefing.

	Rank	Value	% of	Min.	Ave. min	Total min. spent
Top Weather Information Providers	0-1	0-1	flights	spent	spent	on all sources
Public NWS or NOAA site	0.7	0.8	49.8	**13.9**	**6.9**	**19.8**
FSS (standard)	**1.0**	**1.0**	**61.5**	9.1	5.6	
Top Weather Products						**16.6**
METAR	**1.0**	**1.0**	**77.3**	**4.5**	**3.4**	
TAF	**1.0**	**1.0**	**76.5**	**5.3**	**4.0**	
AIRMET / SIGMET	**0.5**	**0.7**	**47.6**	3.7	1.8	
ATIS (Automated Terminal Information System)	**0.4**	**0.5**	**41.4**	2.0	0.8	
charts, Radar (NEXRAD)	**0.5**	**0.6**	**44.2**	3.6	1.6	
FA (Aviation area 18-hr forecast)	**0.5**	**0.5**	**36.1**	3.2	1.2	
Top En-route Weather Information Sources						**7.3**
ATIS	**1.0**	**1.0**	75.6	4.6	3.5	

DISCUSSION AND CONCLUSIONS

Table 5 summarizes the top weather information providers, products, and en-route sources, as rated by the pilots sampled. The first question that comes to mind is whether 16-20 minutes preflight preparation and 7-8 minutes en-route followup are sufficient to prepare for a 4-hour flight into weather challenging to both the pilot's skill and the aircraft's capabilities. If the time is sufficient, a) how efficiently is the time spent, and b) what is the minimum time necessary? Although this study can't address cognitive efficiency, it may be able to address minimums.

Table 6 summarizes the estimated average number of minutes spent on weather briefing for preflight providers and products, and for en-route sources. Minimums, maximums, ranges, and standard deviations are also shown.

Table 6. Estimated average time spent on weather briefings by providers, products, and enroute sources.

	Providers	Products	En-route
Average time spent	19.8	16.6	7.3
Minimum	**3.10**	**3.97**	**0.99**
Maximum	138.5	154.6	92.0
Range	135.4	150.63	91.01
Standard deviation	24.5	23.9	12.9

Conclusion #1 is that, despite the acceptable group averages, given the wide range and large standard deviation, *there seem to be individuals spending as little as 3-4 minutes on preflight weather briefing, and less than one minute on updates, once airborne*. Perhaps these numbers point to a group we should be concerned with, namely those at the short-time end of the distribution. Conclusion #2 is that, while many pilots seem to value and use the modern, sophisticated information providers, *there seems to be a strong, counter-tendency to value and use that which is simplest*. As Table 5 shows, the most popular weather information products and en-route sources sampled here seem to be among the simplest. This has implications for user interface design, certification, and training. It also may reflect a deep problem for some pilots, given the inherently complex nature of weather.

Regarding suggestions for further study, it is recommended that fewer polling variables are needed (specifically *frequency of use* and *Average Minutes Spent*). Future studies should also consider exploring flight duration as a variable, and should explore whether the "low-use/simple-use" pilots described here constitute an at-risk group.

REFERENCES

Miles, M.B., & Huberman, A.M. (1994). Qualitative Data Analysis. Thousand Oaks, CA: Sage.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the efforts of Jerry Ball and of the various site points-of-contact who made data collection possible, Roni Prinzo who was responsible for acquiring and analyzing the AFSS voice tapes and contributing to the manuscript, and Dennis Beringer for data collection and editorial contributions.

Visibility in the Aviation Environment

Michael A. Crognale, Ph.D.
The University of Nevada, Reno

Many general aviation accidents can be partly attributed to failure of the pilot to detect other aircraft and/or conditions of reduced visibility resulting in controlled flight into terrain (CFIT) or even ground based accidents and runway incursions. This present project (currently at the end of its second year) is aimed at better understanding visual limitations imposed by factors in the aviation environment and to provide interactive educational materials with the aim of teaching pilots how to deal with these limitations and to better recognize unsafe visual conditions.

Introduction

General

The present report represents the second annual report for this project due to a late funding date of April 2003 and covers activity from April of 2004 until October of 2005. We will describe below several important results and accomplishments during this period.

Purpose

Each year there are a large number of accidents in general aviation that result in controlled flight into terrain (CFIT) or collision with other aircraft or land based obstructions such as radio towers (Khatwa& Roelen,1996; O'Hare & Owen, 2002; Volpe, 1994). These accidents occur not only when there is continued visual flight into instrument meteorological conditions (IMC), but often times in conditions of clear weather (reviewed by Kraus, 1995; O'Hare & Owen, 2002). The problem of not being able to visually acquire other aircraft and terrain has its roots in several important issues wo of which are considerd here.

1) *Learning to see the target*- Visual detection is an active task rather than a passive one. Efficient search and detection requires that the observer know what to look for, that is approximately where, when, and how it will appear. The solution to these tasks are easily calculated from known relationships. Training is required however for pilots to perform quickly and automatically.

Last year we developed a cockpit aid for training pilots to better judge distance and size of targets. This product was met with great enthusiasm and I am still getting correspondence requesting this product for flight training at flight schools in Civil and military contexts.

We will describe below the current design of additional educational products that should aid the pilot in learning to see other aircraft in the flight environment.

2)*Learning to judge the visual environment*- There are three components to this issue a) the background, b) intervening atmosphere and c) lighting especially "flat-light".

The background against which targets must be detected varies from low contrast, uniform (e.g. clear blue sky) to complex and high contrast (e.g. cityscapes and mottled mountainous terrain). In general, detection is inversely related to scene complexity. In other words, the more complex and higher contrast the background, the harder it is to detect a target on it.

In order to train pilots to judge conditions under which detection may be difficult we must first have a way to characterize the background. We must then model detection on different backgrounds composed of images from the aviation environment.

In addition to research on the effects of backgrounds on detection, we have begun to investigate evolutionary adaptation to the aviation environment. Although it has been argued that most natural images show frequency spectra that fall off in amplitude as 1/f, there is ample evidence that the spectra of many scenes differ from 1/f significantly (e.g. Field & Brady, 1997). Last year we applied sparse coding algorithms to images from the aviation environment (Simoncelli & Olshausen, 2001). This algorithm produces basis functions which are believed to be generated in a similar manner to the receptive fields of visual cortical neurons, that is, by learning from the environment. Such an application provides insight as to the limits of applying our land based visual system to the demands of the aerial environment. Have reported these results last year.

The second and third parts of learning to judge the visual environment (intervening atmosphere and lighting) are concerned largely with weather phenomenon. Whenever there is visible moisture, smoke, or other particulate matter in the air, visibility will be reduced. The visual effects of intervening atmosphere are well modeled by reduction in contrast and a diffusion of the light source. However, these factors can vary independently and have independent effects on the visual system.

While reduction of contrast will reduce the ability to detect outside objects increasingly with distance, light scatter may not. Light scatter may occur well above and below the path of the aircraft such that visibilities are essentially unrestricted yet depth perception and to some degree target detection will suffer greatly. Such conditions occur when flying over snow fields or water and dessert areas with a well diffusing overcast. Because the light is efficiently diffused in all directions, shadows are completely lost and judgment of distance and many target features are greatly disturbed. Pilots have been known to misjudge distance to targets and the ground, the slope of surfaces, and fail to detect large ground features (e.g. mounds of snow or sand) often with disastrous results.

To address the issue of flat light we plan to develop experimental procedures to quantify the degree of diffusion in an environment and to measure behavioral performance in simulated flat light conditions. The results from these experiments will provide input to educational materials described below.

Accomplishments and Results

Simulator

Last year we completed construction of a flight simulator with extended visual display. This year we have made progress towards programming the simulator to provide appropriate backgrounds and weather phenomenon for detection experiments.

Fig. 1 Simulator for detection experiments.

Aviation Images

Last year we collected high quality digital images from the aviation environment over a large portion of the mainland U.S. and around the greater Anchorage area in Alaska. We also analyzed these images using sparse coding algorithms and compared the characteristics of the aviation environment with those of the terrestrial environment and found that they differ in many important respects. The analysis has allowed us to quantify those differences.

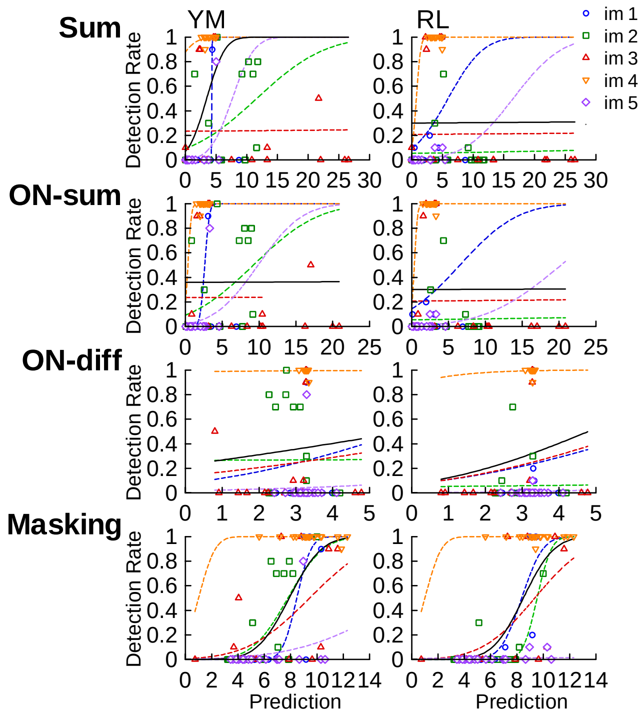


Figure 2. Detection data and model fits for 5 different images from the aviation environment.

This year we have developed three related models of detection based on the sparse coding information and compared them with another model of detection (Ahumada and Beard) as well as actual detection performance data collected in our lab (see fig.2).

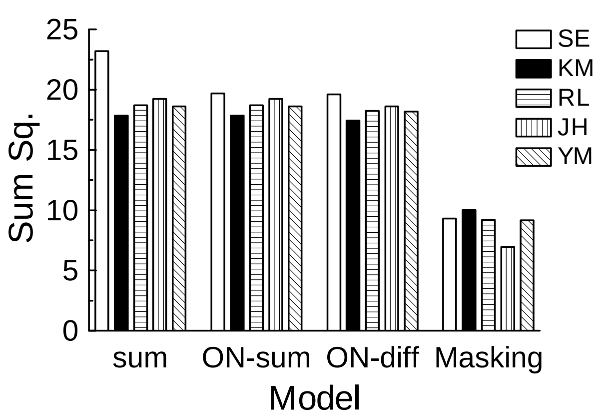


Figure 3. Sum of squared error for fits of four different models of detection with actual detection data for 5 different subjects. The masking model (Ahumada and Beard) appeared to better account for the data.

We found that although the image analysis based on sparse coding was quite useful for quantifying the image characteristics, the models developed using the algorithm did not provide significant advantage over the mathematically simpler Ahumada and Beard model (see figs. 2 & 3)(Mizokami & Crognale, 2005b).

The next phase of this study will be to test the predictions of the models against behavioral detection results obtained in a more realistic aviation setting include distractions and flying tasks provided by the flight simulator.

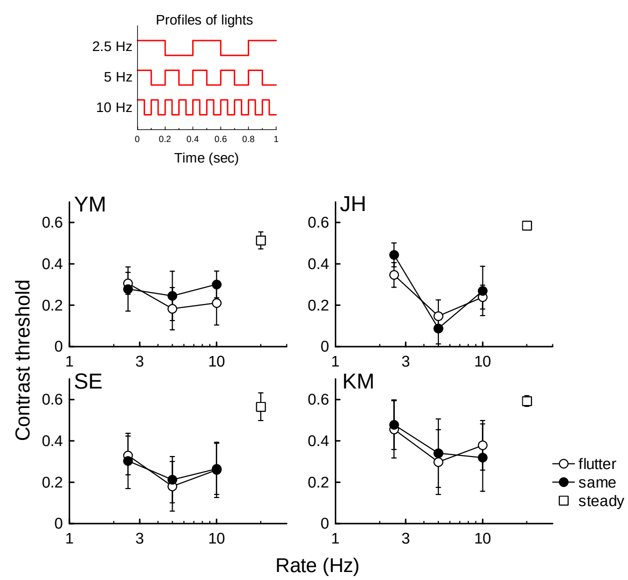
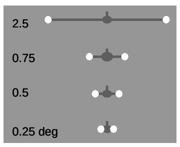


Figure 4. Effects of strobe frequency on detection shown for 4 subjects.

Another practical issue in aviation detection is how to improve aircraft detection through the use of lights. Strobe lights provide a means for improving detection of an aircraft. Some evidence suggests that rate of flash and the percept of apparent motion created by flashing the lights a synchronously might improve detection over a traditional synchronous paired strobe flash pattern. We tested this directly in a series of experiments that required detection of flashed lights on noise backgrounds that emulated that found in the aviation environment. We found an obvious improvement in detection for flashing strobes vs. steady lights but little effect of degree of synchrony, rate of flash and distance between strobes over a range relevant for aircraft detection (see figs. 4-7) (Mizokami and Crognale, 2005a).



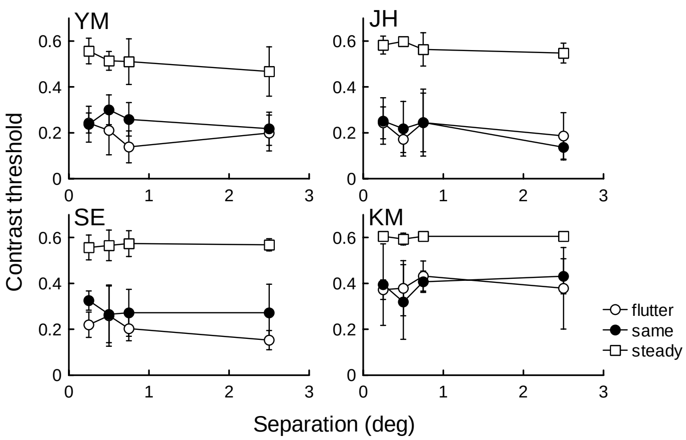
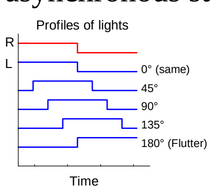


Figure 5. Effects of separation distance on detection of steady, and synchronous and asynchronous strobes.



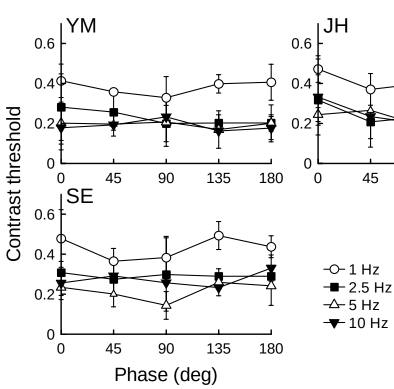
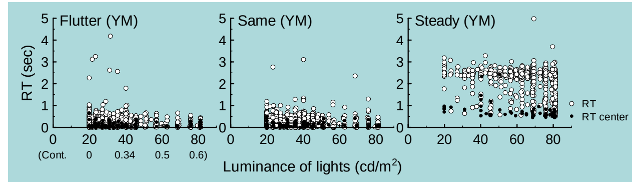


Figure 6. Effects of phase of synchrony on detection for



3 subjects. Phase has little effect.

Figure 7. Effects of mode of flash (asynchronous flash, synchronous flash and steady) on reaction time to detection.

Learning to see

Last year we developed a simple reference card for use in the cockpit (see appendix). This card illustrates the apparent sizes of typical small airplanes (e.g. Cessna 172) and airliners (e.g. Airbus A-320) at different distances from 2 miles to ½ mile. This card can be used by the pilot to estimate the approximate size of a known but undetected target. Feedback on the use of this card has been quite positive and we will continue to provide it to pilots as requested.

The first part of the program introduces the concept of visibility in the context of the aviation environment. The second part introduces 4 problem areas:1) learning to see; 2) VFR fight into IMC; 3) background masking; and 4) flat light. The third part will be interactive training in two main areas 1) learning to see other aircraft and 2) learning to evaluate the visual environment. The first part will cover judgments of distance, direction, altitude, flight path and orientation. The second part will cover judgments of background masking effects, atmospheric haze , VFR into IMC, and flat light recognition.

We have completed a preliminary version of the part of the program that trains pilots how to judge the appearance and elevation of aircraft traffic given the distance, direction of flight, and altitude from a simulated traffic call. The trainee is also given an altimeter readout and a directional gyro readout in order to provide information to compute relative orientation and altitude. The trainee's task is to pick the visual scenario that matches the traffic call, out of four possible scenarios that appear on the screen simultaneously. The trainee is also provided feedback to improve learning.

The final main deliverable product should be available by the end of the 3-year funding period (March 31, 2006).

References

Ahumada A.J (1996) Simplified Vision Models for Image Quality Assessment. SID International Digest of Technical Papers, Volume XXVII, May, 397-400.

Ahumada A.J. & Beard B.L. (1997) Image Discrimination Models Predict Detection in Fixed but not Random Noise. JOSA A, 14, 2471-2476.

Ahumada A.J. & Beard B.L. (1998) A simple Vision Model for Inhomogeneous Image Quality Assessment. SID Digest of Technical Papers XXIX, paper 40.1.

Field D.J. & Brady N. (1997) Visual Sensitivity, Blur and the sources of Variability in the Amplitude Spectra of Natural Scene. Vision Research 37, 3367-3383.

Khatwa R. & Roelen A.L.C. (1996) An Analysis of Controlled-flight-into-terrain (CFIT) Accidents of Commercial Operators 1988 through 1994, Flight Safety Foundation, Pilot Safety Digest, Special Issue, April-May.

Krause R.R (1995) Avoiding Mid-Air Collisions. S.S. Krause, Tab Books.

Mizokami Y. & Crognale M.A. (2005a) Detection of dual flashing lights. *Annual meeting for the Vision Sciences Society*, Sarasota, FL.

Mizokami Y. & Crognale M.A. (2005b) Detection model predictions for aircraft targets on natural backgrounds. European Conference on Visual Perception.

O'Hare, D. & Owen D (2002) Cross-country VFR crashes: pilot and contextual factors.. Aviation Space and Environmental Medicine (73) 363-366.

Olshausen, B.A., & Field D.J. (1996). Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. Nature, 381, 607-609.

Olshausen, B.A., & Field D.J. (1997). Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vision Research, 37, 3311-3325.

Rohaly A.M., Ahumada, A.J., Watson A.B. (1997) Object Detection in Natural Backgrounds predicted by Discrimination Performance and Models. Vision Research 37, 3225-3235.

Simoncelli, E.P & Olshausen B.A. (2001) Neural Image Statistics and Neural Representation, E.P. Annual Reviews of Neuroscience, 24,1193-1216.

Volpe (1994) Controlled Flight Into Terrain (CFIT) Report, 1983-1994