Office of the Chief Scientist for Human Factors

Human Factors General Aviation

Program Review
FY03

Dr. William K. Krebs, General Aviation Human Factors Program Manager

Federal Aviation Administration
AAR-100 (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 (AAR-100) 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, 2002 and December 31st, 2003. 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 AAR-100 general aviation program manager on the quality of the research program. Basically, this document is a means of holding each group (sponsor, investigator, AAR-100 program manager) accountable to ensure that the program is successful.

In FY03, the general aviation research program distributed $625,000 contract dollars to seven performing organizations. In addition, some of these projects received supplemental support from the Civil Aerospace Medical Institute, Oklahoma City, OK. These projects are described in Appendix I and the requirements that are mapped to these projects are located in Appendix II.

The FY04 funded projects ($437,500 contract dollars) and the proposed FY05 (estimated $437,500 contract dollars) and FY06 projects (estimated $437,500 contract dollars).

To view projects, pages 4-28

To view requirements, pages 29-67

Address questions or comments to:

William K. Krebs, Ph.D.

Appendix I

Human Factors General Aviation

FY03 Project Summaries

Human Error and General Aviation Accidents: A Comprehensive, Fine-Grained Analysis Using HFACS

Scott A. Shappell, Ph.D.
FAA/Civil Aerospace Medical Institute

Douglas A. Wiegmann, Ph.D.
University of Illinois at Urbana-Champaign

A large effort has been expended over the last several decades to lower the military and commercial aviation accident rates. Unfortunately, until recently, a similar effort has not occurred within the general aviation (GA) community even though the total number of accidents is considerably greater. As part of the FAA's endeavor to better understand the etiology of GA accidents we previously analyzed eleven years (1990-2000) of GA accidents using the Human Factors Analysis and Classification System (HFACS). The findings, though significant, spawned additional questions regarding the nature of aircrew error associated with GA accidents. For instance, how often is each error type the "initiating" error in the causal chain of events and what are the exact types of errors committed within each error category? This brief report details the efforts made by the University of Illinois and the FAA Civil Aerospace Medical Institute to address these questions in FY 2003.

INTRODUCTION

It is generally accepted that like most accidents, those in aviation do not happen in isolation. Rather, they are often the result of a chain of events often culminating with the unsafe acts of aircrew. Indeed, from Heinrich's (Heinrich, Peterson, & Roos, 1931) axioms of industrial safety, to Bird's (1974) "Domino theory" and Reason's (1990) "Swiss cheese" model of human error, a sequential theory of accident causation has been consistently embraced by most in the field of human error. Particularly useful in this regard has been Reason's (1990) description of active and latent failures within the context of his "Swiss cheese" model of human error.

In his model, Reason describes four levels of human failure, each one influencing the next. Included were: 1) Organizational influences, 2) Unsafe supervision, 3) Preconditions for unsafe acts, and 4) the Unsafe acts of operators. Unfortunately, while Reason's seminal work forever altered the way aviation and other accident investigators view human error; it did not provide the level of detail necessary to apply it in the real world.

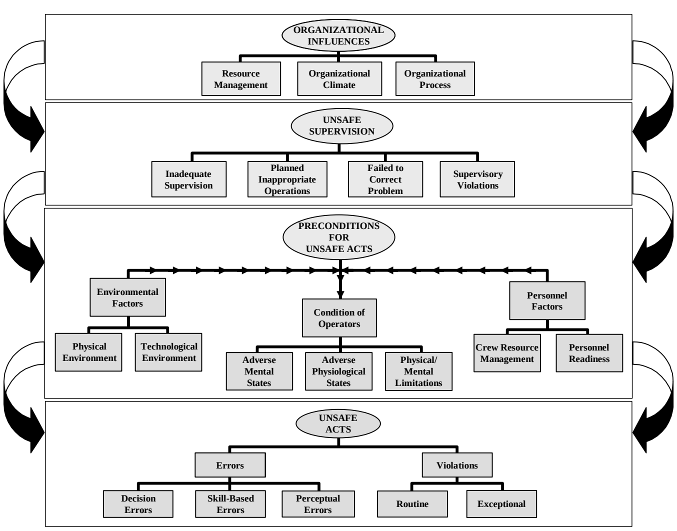

Figure 1. The HFACS framework.

It wasn't until Shappell and Wiegmann, (2000, 2001) developed a comprehensive human error framework - the Human Factors Analysis and Classification System (HFACS); that Reason's ideas were folded into the applied setting. The HFACS framework includes 19 causal categories within Reason's (1990) four levels of human failure (Figure 1). Unfortunately, a complete description of all 19 causal categories is beyond the scope of this brief report. It is however, available elsewhere (Wiegmann and Shappell, 2003).

HFACS

Particularly germane to any examination of GA accident data are the unsafe acts of aircrew – all the while keeping in mind that data from the preconditions for unsafe acts, and in some instances unsafe supervision and organizational influences, are important as well. For that reason, we will briefly describe the causal categories associated with the unsafe acts of GA aircrew.

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 surprising, 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.

Errors

Within HFACS, the category of errors was expanded to include three basic error types (decision, skill-based, and perceptual errors).

Decision Errors. Decision-making and decision errors have been studied, debated, and reported extensively in the literature. In general however, decision errors can be grouped into one of three categories: procedural errors, poor choices, and problem solving errors. Procedural decision errors (Orasanu, 1993) or rule-based mistakes as referred to by Rasmussen, (1982) occur during highly structured tasks of the sorts, if X, then do Y. Aviation is highly structured, and consequently, much of pilot decision-making is procedural. That is, there are very explicit procedures to be performed at virtually all phases of flight. Unfortunately, on occasion these procedures are either misapplied or inappropriate for the circumstances often culminating in an accident.

However, even in aviation, not all situations have corresponding procedures to manage them. Therefore, many situations require that a choice be made among multiple response options. This is particularly true when there is insufficient experience, time, or other outside pressures that may preclude a correct decision. Put simply, sometimes we chose well, and sometimes we do not. The resultant choice decision errors (Orasanu, 1993), or knowledge-based mistakes (Rasmussen, 1982), have been of particular interest to aviation psychologists over the last several decades.

Finally, there are instances when a problem is not well understood, and formal procedures and response options are not available. In effect, aircrew find themselves where they have not been before and textbook answers are nowhere to be found. It is during these times that the invention of a novel solution is required. Unfortunately, individuals in these situations must resort to slow and effortful reasoning processes; a luxury rarely afforded in an aviation emergency – particularly in general aviation.

Skill-based Errors. Skill-based behavior within the context of aviation is best described as "stick-and-rudder" and 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. In fact, attention failures have been linked to many skill-based errors such as the breakdown in visual scan patterns, inadvertent activation of controls, and the misordering of steps in procedures. Likewise, memory failures such as omitted items in a checklist, place losing, or forgotten intentions have adversely impacted the unsuspecting aircrew.

Equally compelling yet not always considered by investigators is the manner or technique one uses when flying an aircraft. Regardless of one's training, experience, and educational background, pilots vary greatly in the way in which they control their aircraft. Arguably, such techniques are as much an overt expression of ones personality as they are a factor of innate ability and aptitude. More important however, these techniques can interfere with the safety of flight or may exacerbate seemingly minor emergencies experienced in the air.

Perceptual Errors. While, decision and skill-based errors have dominated most accident databases and have therefore been included in most error frameworks, perceptual errors have received comparatively less attention. No less important, 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. Faced with acting on inadequate information, aircrew run the risk of misjudging distances, altitude, and decent rates, as well as responding incorrectly to a variety of visual/vestibular illusions.

It is important to note, however, that it is not the illusion or disorientation that is classified as a perceptual error. Rather, it is the pilot's erroneous response to the illusion or disorientation that is captured here. For example, many pilots have experienced spatial disorientation when flying in IMC. In instances such as these, pilots are taught to rely on their primary instruments, rather than their senses when controlling the aircraft. Still, some pilots fail to monitor their instruments when flying in adverse weather or at night, choosing instead to fly using fallible cues from their senses. Tragically, many of these aircrew and others who have been fooled by visual/vestibular illusions have wound up on the wrong end of the accident investigation.

Violations

By definition, errors occur while aircrews are behaving within the rules and regulations implemented by an organization. In contrast, violations represent the willful disregard for the rules and regulations that govern safe flight and, fortunately, occur much less frequently (Shappell and Wiegmann, 1995).

Routine Violations. While there are many ways to distinguish between types of violations, two distinct forms have been identified, based on their etiology. The first, routine violations, tend to be habitual by nature and are often tolerated by the governing authority (Reason, 1990). Consider, for example, the individual who drives consistently 5-10 mph faster than allowed by law or someone who routinely flies in marginal weather when authorized for VMC only. While both certainly violate governing regulations, many drivers or pilots do the same thing. Furthermore, people who drive 64 mph in a 55-mph zone, almost always drive 64 in a 55-mph zone. That is, they *routinely* violate the speed limit.

Often referred to as "bending the rules," these violations are often tolerated and, in effect, sanctioned by authority (i.e., you're not likely to get a traffic citation until you exceed the posted speed limit by more than 10 mph). If, however, local authorities started handing out traffic citations for exceeding the speed limit on the highway by 9 mph or less, then it is less likely that individuals would violate the rules. By definition then, if a routine violation is identified, investigators must look further up the causal chain to identify those individuals in authority who are not enforcing the rules.

Exceptional Violations. In contrast, exceptional violations appear as isolated departures from authority, not necessarily characteristic of an individual's behavior nor condoned by management (Reason, 1990). For example, an isolated instance of driving 105 mph in a 55 mph zone is considered an exceptional violation. Likewise, flying under a bridge or engaging in other particularly dangerous and prohibited maneuvers would constitute an exceptional violation. However, it is important to note that, while most exceptional violations are indefensible, they are not considered exceptional because of their extreme nature. Rather, they are considered exceptional because they are neither typical of the individual nor condoned by authority. Unfortunately, the unexpected nature of exceptional violations makes them particularly difficult to predict and problematic for organizations to manage.

Previous Findings

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.

When the GA accidents between 1990-2000 were examined using the HFACS framework; several heretofore unknown facts regarding GA aviation safety were revealed (Figure 2). For instance, it appears that safety efforts over the last 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). Given that most of these skill-based errors were technique (stick-and-rudder) errors, it would seem to indicate that there may be a problem associated with current training and/or pilot currency/proficiency.

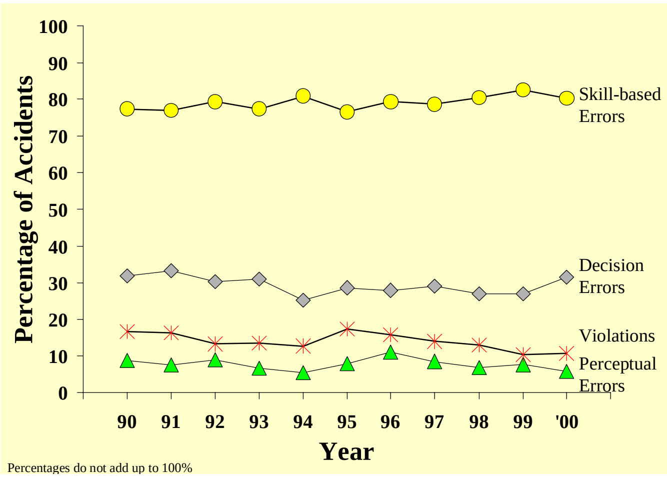

Figure 2. Overall percentage of GA accidents associated with at least one instance of a given unsafe act.

Furthermore, when the data are separated into fatal (Figure 3) and non-fatal (Figure 4) accidents, clear differences in the pattern of human error were noted. For example, while skill-based errors remained the predominant error form observed during both fatal and non-fatal accidents, violations of the rules were much more likely to occur during fatal than non-fatal GA accidents. The data also suggest that if a GA pilot elects to continue into IMC when he/she is VFR only (the predominant violation observed in the data), they are over 3 times more likely to die or kill someone else.

Although there was some variation, there were no significant differences observed between fatal and non-fatal accidents for decision or perceptual errors. That is, decision errors were observed in roughly 30% of the fatal and non-fatal accidents examined, while perceptual errors were associated with less than 10% of fatal and non-fatal accidents.

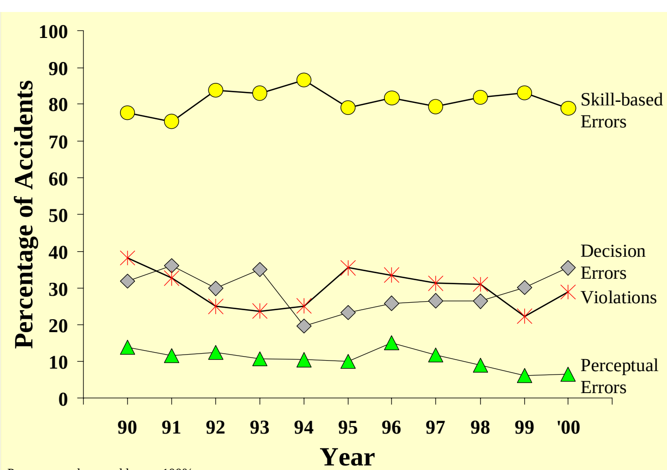

Figure 3. HFACS analysis of fatal GA accidents.

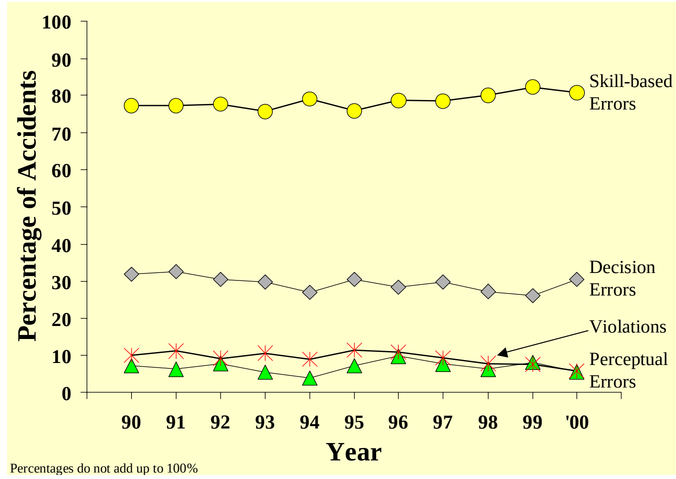

Figure 4. HFACS analysis of non-fatal GA accidents.

FY03 Research Effort

Key members of the FAA (e.g., AFS-800) and several committees chartered to address GA 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 the HFACS analyses. However, these individuals and committees have requested that additional analyses be done to answer specific questions regarding the nature of the human errors identified within the context of GA. For instance:

- How important is each error type? That is, how often is each error type the "primary" cause of an accident? For example, 80% of accidents might be associated with skill-based errors; but how often are they the "initiating" error or simply the "consequence" of another type of error, such as decision errors?
- What are the exact types of errors committed within each error category? In other words, how often do skill-based errors involve stick-and-rudder errors, verses attention failures (slips) or memory failures (lapses) and what are those errors specifically?

Answers to these questions were not available in the database, as it currently existed. Therefore, additional fine-grained analyses of the specific human error categories within HFACS were needed to answer these and other questions that have arisen, and to target problem areas within GA for future interventions. A new requirement was therefore initiated in 2002 to address these questions. CAMI and the University of Illinois are now midway through the second year of a three-year effort to perform a 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.

METHOD

Data

As with the previous studies (above), GA accident data from calendar years 1990-2000 was obtained from databases maintained by the NTSB and the FAA's National Aviation Safety Data Analysis Center (NASDAC). In total, 20,797 GA accidents were extracted for analysis. These so-called "GA" accidents actually included a variety of aircraft being flown under several different operating rules: 1) 14 CFR Part 91 – Civil aircraft other than moored balloons, kites, unmanned rockets, and unmanned free balloons; 2) 14 CFR Part 91F – Large and turbine-powered multiengine airplanes; 3) 14 CFR Part 103 – Ultralight vehicles; 4) 14 CFR Part 125 – Airplanes with seating capacity of 20 or more passengers or a maximum payload capacity of 6,000 pounds or more; 5) 14 CFR Part 133 – Rotorcraft external-load operations; 6) 14 CFR Part 137 – Agricultural aircraft operations. In addition, the database contained several accidents involving public use aircraft (i.e., law enforcement, state owned aircraft, etc.) and a few midair accidents involving military aircraft.

Although all 20,797 accidents obtained can be found within the NTSB under the heading of "general aviation," we were only interested in those accidents involving aircraft operating under 14 CFR Part 91. After all, it is difficult to envision that large commercial aircraft being ferried from one airport to the next (operating under 14 CFR Part 91F) or aircraft being used to spread chemicals on a field (operating under 14 CFR Part 137) can be equated with small private aircraft being flown for personal or recreational purposes (operating under 14 CFR Part 91). This left us with 19,147 accidents in the database.

For this analysis we were primarily concerned with powered aircraft and therefore conducted another reduction of the data to include only accidents involving powered fixed-wing aircraft (i.e., no gliders, ultra-lights, balloons, or blimps), helicopters, and gyrocopters. The remaining 18,531 accidents were then examined for aircrew-related causal factors. Since we were only interested in those involving aircrew error, not those accidents that were purely mechanical in nature or those solely attributable to other human involvement, a final reduction of the data was conducted. Note, this does not mean that mechanical failures or other sources of human error did not exist in the final database, only that some form of aircrew error was also involved in each of the accidents included. In the end, 14,631 accidents were included in the database and submitted to further analyses using the HFACS framework.

<u>Causal Factor Classification using HFACS</u>

Five 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 five were certified flight instructors with a minimum of 1,000 flight hours in GA aircraft (mean = 3,530 flight hours) as of June 1999 when the study began. After training, the five GA pilot-raters were randomly assigned accidents so at least 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 (29,676 agreements; 4,474 disagreements).

RESULTS

Unlike our previous studies where we were interested in the percentage of accidents associated with at least one instance of a given unsafe act, our focus this FY has been on identifying the seminal (precipitating) aircrew unsafe act. That is, what percentage of the time are skill-based errors, decision errors, perceptual errors, and violations the <u>first</u> unsafe act committed by the aircrew in the chain of events leading to an accident. The results were very similar to those seen in our previous studies.

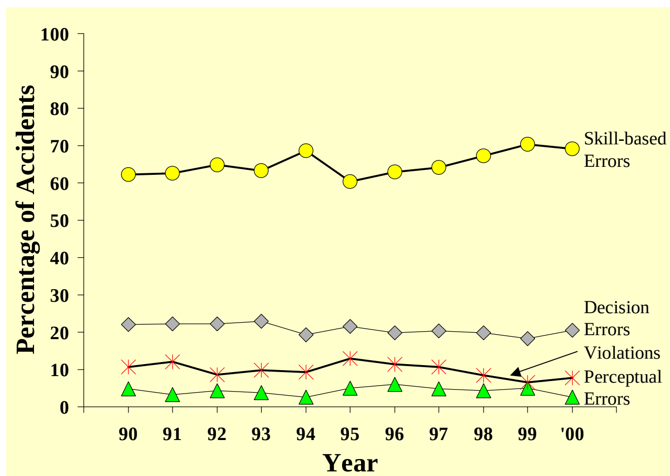

Figure 5. Seminal HFACS analysis of GA accidents.

An examination of the overall seminal HFACS analysis (Figure 5) revealed that, as before, skill-based errors were the most frequently cited seminal unsafe act by an almost 3 to 1 margin. These were followed by decision errors, violations, and perceptual errors in that order. Note that unlike the data from the previous studies, the percentages here do add up to 100% since there is only one seminal (precipitating) error per accident.

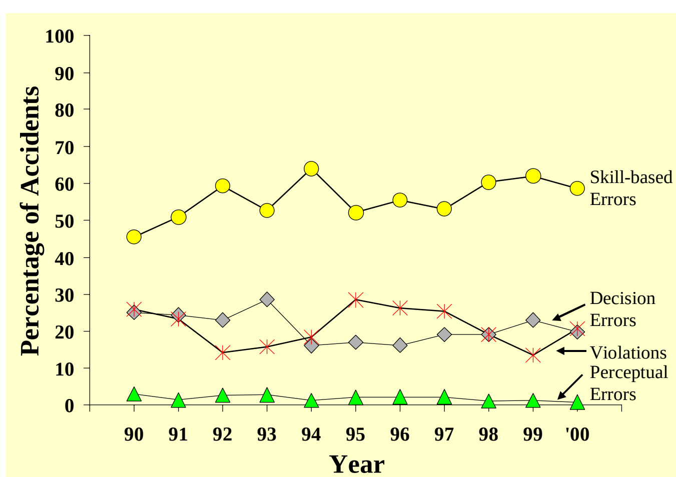

Figure 6. Seminal HFACS analysis of fatal GA accidents.

Even when the data are analyzed separately for fatal (Figure 6) and non-fatal GA accidents (Figure 7), the pattern of errors remained essentially unchanged. That is, skill-based errors were the most frequently cited seminal unsafe act. The only notable difference was that considerably more violations were seminal in the chain of events leading up to a fatal accident when compared to non-fatal accidents.

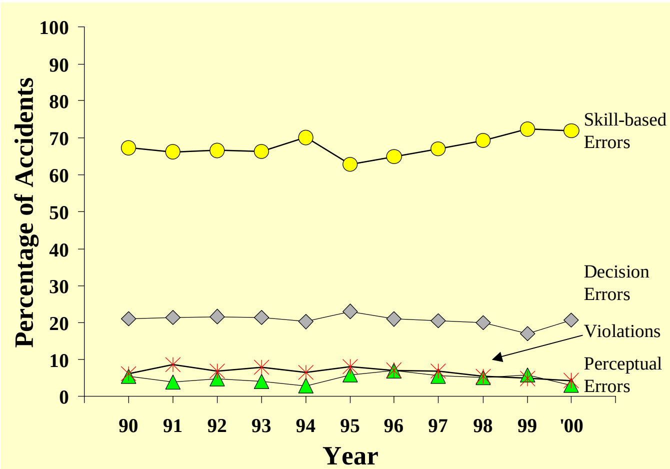

Figure 7. Seminal HFACS analysis of non-fatal GA accidents.

DISCUSSION

It would appear from our fine-grained analyses that it doesn't matter whether one examines the percentage of accidents associated with at least one instance of a given unsafe act or the seminal unsafe act, the pattern of human error observed among GA accidents remains essentially the same. That is, skill-based errors are consistently the most common error leading to a GA accident and in most cases is the seminal error form as well. Furthermore, when violations are associated with GA accidents they are much more likely to result in a fatality than if a violation is not committed. It is also noteworthy that while a great deal of effort has been expended to inform pilots of the hazards of spatial disorientation and visual illusions, it does not appear to have been done in vane since perceptual errors are the least common among all four categories of unsafe acts.

With the issue of seminal (precipitating) causal factors seemingly answered (i.e., the pattern of human error did not change appreciably from that previously reported looking at all aircrew errors), our work can now turn toward an examination of the specific types of skill-based errors, decision errors, perceptual errors, and violations that are most predominant among the unsafe acts. To give the reader a sense of what that analysis will look like, a preliminary analysis of the seminal skill-based errors was conducted and the results presented in Table 1.

Table 1. Top Five Seminal Skill-based Errors

Directional control	1357
Airspeed	1045
Compensation for winds	867
Aircraft control	809
Visual lookout	365

It is clear from the table that the top five types of skill-based errors all involve technique (stick-and-rudder/basic flight skills) errors rather than errors due to failures of attention or memory. This is important since it suggests that improved or additional training (both *ab initio* and recurrent) is needed to prevent or mitigate these types of errors.

The good news is that AFS-800 has recently introduced the FAA/Industry Training Standards (FITS) program aimed at improving GA flight training. While the program is currently focusing on "personal or professionally flown single-pilot aircraft for transportation with new technologies," (Glista, 2003) there is no reason to believe that FITS will not benefit the light-sport and recreational pilots as well. Furthermore, data from the HFACS analysis will provide valuable information for the FAA and other civilian organizations as they develop data-driven intervention and prevention strategies for the GA community.

REFERENCES

Bird, F. (1974). *Management guide to loss control.* Atlanta, GA: Institute Press.

Glista, T. (2003). FAA/Industry Training Standards (FITS): Times(and training requirements are a changing). *FAA Aviation News*, *42*(4), 1-6.

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

Orasanu, J. M. (1993). Decision-making in the cockpit. In E. L. Wiener, B. G. Kanki, and R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 137-172). San Diego, CA: Academic Press.

Rasmussen, J. (1982). Human errors: A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4*, 311-333.

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

Shappell, S. & Wiegmann, D. (1997). A human error approach to accident investigation: The Taxonomy of Unsafe Operations. *The International Journal of Aviation Psychology*, *7*, 269-291.

Shappell, S. & Wiegmann, D. (2000). *The Human Factors Analysis and Classification System (HFACS).* Federal Aviation Administration, Office of Aviation Medicine Report No. DOT/FAA/AM-00/7. Office of Aviation Medicine: Washington, DC.

Shappell, S. & Wiegmann, D. (2001). Applying Reason: the human factors analysis and classification system (HFACS). *Human Factors and Aerospace Safety, 1*, 59-86.

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

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

THE EFFECTIVENESS OF A PERSONAL COMPUTER AVIATION TRAINING DEVICE (PCATD), A FLIGHT TRAINING DEVICE (FTD), AND AN AIRPLANE IN CONDUCTING INSTRUMENT PROFICIENCY CHECKS

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

This project seeks to evaluate the effectiveness and reliability of a personal computer aviation training device (PCATD) and a flight training device (FTD) in conducting an IPC. The study will compare the performance of pilots receiving an IPC in a PCATD, in a FTD or in an airplane (IPC #1) with performance on an IPC in an airplane (IPC #2). This comparison between a PCATD and an airplane will investigate the effectiveness of the PCATD in administering an IPC. Currently, the PCATD is not approved for IPCs. The comparison between a Frasca and the airplane will determine whether the current rule to permit IPCs in a FTD is warranted. Finally, the performance of pilots receiving IPC #1 in an airplane will be compared with IPC #2 in an airplane with a second CFII. This comparison will permit the determination of the reliability of IPCs conducted in an airplane.

INTRODUCTON

To maintain instrument currency, instrument pilots must meet the recency of experience requirements of FAR 61.57(c) or (d) every six months. The recency of experience requirements may be conducted in an airplane or simulated in an approved flight training device (FTD). If an instrument pilot fails to meet recency of experience requirements within a 12-month period, an instrument proficiency check (IPC) must be accomplished with a certified flight instructor, instrument (CFII) to regain instrument currency.

Taylor, Lintern, Hulin, Talleur, Emanuel, and Phillips (1996, 1999) conducted a study to determine the extent to which a personal computer aviation training device (PCATD) can be used to develop specific instrument skills that are taught in instrument flight training and to determine the transfer of these skills to the aircraft. This in turn led to an additional study by the Institute of Aviation of the University of Illinois at Urbana-Champaign (UIUC) to determine the effectiveness of PCATDs for maintaining instrument currency (Taylor, Talleur, Bradshaw, Emanuel, Rantanen, Hulin and Lintern, 2001; Talleur, Taylor, Emanuel, Rantanen, and Bradshaw, in press). In the latter study, a total of 106 instrument current pilots were divided in four groups. The pilots in each group received an instrument proficiency check (IPC #1). During a six-month period following IPC #1, the pilots in three groups received recurrent training in a PCATD, a Frasca flight training device (FTD), or an airplane, respectively. The fourth (control) group received no training during the six-month period. After this time, the pilots in each group flew an instrument proficiency check (IPC #2). The comparison of IPC #1 and IPC #2 indicated that both the PCATD and the Frasca FTD were more effective in maintaining instrument proficiency when compared to the control group and at least as effective as the airplane. The study also found that of 106 instrument current pilots, only 45 (42.5%) were able to pass IPC #1. Of the group who received an IPC in a Frasca FTD to regain currency, only 22 of 59 were able to subsequently able to pass IPC #1 in an airplane. This study established the effectiveness of PCATDs for use in instrument currency training. However, the question of whether PCATDs are effective for administering the IPC has not been demonstrated. Based on the data above a question concerning the effectiveness of the Frasca FTD in administrating an IPC also arises.

The purpose of the present study is to compare the performance of pilots receiving an IPC in a PCATD, a FTD or an airplane (IPC #1) with their performance in an airplane (IPC #2). The comparison of performance in a PCATD to that in an airplane investigates the effectiveness of the PCATD as a device in which to administer an IPC. Currently, the PCATD is not approved to administer IPCs. The comparison of performance in a FTD with performance in an airplane will help determine whether the current rule to permit IPCs in a FTD is warranted. Finally, the comparison of performance of pilots receiving IPC #1 in an airplane and IPC #2 in an airplane with a second CFII will permit the determination of the reliability of IPCs conducted in an airplane.

METHOD

Participants

In the initial proposal a total of 105 pilots (35 in each group) were scheduled to participate in the study. Due to funding reductions in the third year funding, the number of pilots in the study was reduced to a total of 75 pilots (25 subjects in each group; FTD, PCATD and airplane). The participating pilots fall into one of four categories of instrument currency: (1) instrument current, (2) within one year of currency, (3) outside of one year of currency but within two years of currency, and (4) outside two years but within five years of currency. All participants will receive a familiarization flight and a review of the systems and instrumentation in the FTD, the PCATD and the airplane prior to being assigned to an experimental group. A randomization process is being used to balance the order of the familiarization flights. Following the familiarization flights, subjects will be assigned to one of the three groups (FTD, PCATD and Airplane) with a constraint that the currency categories are balanced among the groups. (See Table 1)

Table 1
Randomization Schedule (PCATD = P; Frasca = F; Airplane = A

Replications:

1	2	3	4	5	6
PFA	FAP	APF	PAF	FPA	AFP
FAP	APF	PAF	FPA	AFP	PFA
APF	PAF	FPA	AFP	PFA	FAP
PAF	FPA	AFP	PFA	FAP	APF
FPA	AFP	PFA	FAP	APF	PAF
AFP	PFA	FAP	APF	PAF	

Equipment

Two FAA-approved Elite PCATDs and one FAA-approved Frasca 141 FTD with a generic single-engine, fixed gear, fixed-pitch propeller performance model are being used in the study. Data output and recording systems have been developed for the PCATD and for the Frasca for development and analysis of objective pilot performance measures. The FTD is approved for instrument training towards the instrument rating, instrument recency of experience training, and IPCs as well as for administering part of the instrument rating flight test. Two 180 hp Beechcraft Sundowner aircraft (BE-C23) which have a single engine, fixed-pitch propeller, and fixed undercarriage are being used as aircraft for IPC #1 and IPC #2. These aircraft are equipped with flight data recorders (FDRs) developed at UIUC (Lendrum et al., 2000) for recording of data for objective pilot performance measures (Rantanen & Talleur, 2001).

Procedure

Following the familiarization flights all 75 pilots receive a baseline IPC flight in either the FTD, PCATD or an airplane (IPC #1) according to the group they are assigned. IPC #1 is flown with a certified flight instructor, instrument (CFII) who acts both as a flight instructor and as an experimental observer. Then all subjects are given a second IPC in the airplane (IPC #2) with a second CFII. The participants are required to refrain from instrument flight following IPC #1 until IPC #2 is completed. They must also agree not to use a PCATD or a FTD for instrument training during this period. A limited number of pilots who were more than two years out currency received training an average of six hours training equally distributed among the FTD, PCATD and airplane to prepare them for the IPC. This procedure was discontinued after the second year to reduce expenses. Table 2 depicts the experimental design.

Table 2.
Experimental Design

GROUP	Fam. Flight	Initial IPC flight (IPC#1)	Final IPC flight (IPC#2)
Airplane	In aiplane	IPC flight in Sundowner	IPC flight in Sundowner
Frasca	In Frasca	IPC flight in Frasca	IPC flight in Sundowner
PCATD	In Elite	IPC flight in Elite	IPC flight in Sundowner

The IPC is a standardized test of the instrument pilot's instrument skills. The types of maneuvers, as well as completion standards for an IPC, are listed in the instrument rating practical test standards (PTS) (U.S. Department of Transportation, 1998). A flight scenario that follows the current guidelines for the flight maneuvers required by the PTS is used for the IPC. This scenario is used to collect baseline data and to establish the initial level of proficiency for each subject who participants in the project.

The IPC #1 flight contains six maneuvers (VOR approach, holding pattern, steep turns, unusual altitude recovery, ILS approach and a partial-panel non-precision approach). ATC communication procedures are also scored. The CFIIs for the IPC #1 flight use a form that was designed to facilitate the collection of three types of data (Phillips, Taylor, Lintern, Hulin, Emanuel, & Talleur, 1995). First, within each maneuver there are up to 24 variables (e.g., altitude, airspeed) that are scored as pass/fail indicating whether performance on those variables met PTS requirements. Second, the flight instructor judges whether the overall performance of the each maneuver was pass/fail. Third, the CFII records if the overall performance of the subject met the PTS for the IPC. The instructors who administer the IPC #1 flight have been standardized on the scenario to be flown and the scoring procedure.

After a period not to exceed two weeks, all subjects fly a final IPC (IPC #2) in the aircraft to assess instrument proficiency. IPC #2 is conducted by a different CFII than IPC #1 to eliminate experimenter bias. The CFII for IPC #2 is blind to both the group to which the subject belongs and to the subject's performance on IPC #1. In terms of maneuvers, IPC #2 is identical to IPC #1. This final session contains all required maneuvers that a pilot must satisfactorily complete in order to receive an endorsement of instrument proficiency. Completion of IPC #2 marks the end of a subject's involvement in the experiment.

PRELIMINARY RESULTS

At present, 54 of 75 of intended pilots (72%) have completed IPC #1 and 51 of the 75 pilots (68%) have completed the study. The pass/ fail rates by group for IPC #1 and IPC #2 are shown in Table 3.

Table 3.
Pass/Fail rates by group

IPC#1					
Group	N	Pass	(%)	Fail	(%)
Aircraft	18	4	(22)	14	(78)
FTD	19	5	(26)	14	(74)
PCATD	17	3	(18)	14	(82)
Total	54	11	(20)	42	(78)

IPC#2					
Group	N	Pass	(%)	Fail	(%)
Aircraft	17	8	(47)	9	(53)
FTD	17	7	(41)	10	(59)
PCATD	17	10	(59)	7	(41)
Total	51	25	(49)	26	(51)

A total of 42 of 54 pilots failed IPC #1 (78%) and a total of 26 of 51 pilots failed IPC #2 (51%). The percentages of pilots in each of the three groups who failed IPC #1are as follows: for the Airplane group, 78%, for the FTD group 74% and for the PCATD group 82%. The number of participants who have completed IPC 1 is not sufficient to compute statistical analyses.

The pass/fail rates for IPC #2 in the airplane show fewer failures for each group and for the total when compared to the pass/fail rates for IPC #1. Of the 51 pilots who have taken IPC #2, twenty-five passed (49%) and 26 failed (51%). The failure rate by group was 53% for the Airplane group, 59% for the FTD group and 41% for the PCATD group.

The pass/fail rates by currency status are shown in Table 4. A total of 37 current pilots took IPC #1 and 8 passed (22%) while 29 failed (78%). A total of 35 current pilots have taken IPC #2 and 15 passed (43%) while 20 failed (57%).

A matrix that shows IPC #1 and IPC #2 pass/ fail rates is presented in Table 5. The preliminary data show that 20 pilots who failed IPC#1 passed IPC#2, 18 failed both IPC#1 and IPC#2, 4 passed both IPC#1 and IPC#2 and 9 failed IPC#2 after passing IPC #1.

Table 4.
Pass/Fail rates by currency

IPC #1					
Currency	N	Pas s	(%)	Fail	(%)
Current	37	8	(22)	29	(78)
Within 1 year	6	2	(33)	4	(67)
Within 1-2 years	--	--	--	--	--
2-5 years (Frasca)	5	1	(20)	4	(80)
2-5 years (PCATD)	5	1	(20)	4	(80)

IPC #2					
Currency	N	Pas s	%)	Fail	(%)
Current	35	15	(43)	20	(57)
Within 1 year	6	5	(83)	1	(17)
Within 1-2 years	--	--	--	--	--
2-5 years (Frasca)	5	1	(20)	4	(80)
2-5 years (PCATD)	5	4	(80)	1	(20)

Table 5.
IPC #1 vs. IPC #2 Pass/Fail

		IPC#2		
		Pass	Fail	Tota l
IPC#1	Pass	4	9	13
	Fail	20	18	38
	Total	24	27	51

DISCUSSION

The Federal Aviation Administration permits the use of flight training devices in general aviation training and education. In 1997 the FAA published an advisory circular concerned with the qualification and approval of PCATDs (U.S. Department of Transportation, 1997). The advisory circular permits the use of PCATDs in instrument training programs conducted under FAR Part 61 and Part 141 and authorizes the use of a PCATD to be substituted for 10 of the 15 hours authorized for an approved flight training device (FTD). The advisory circular did not authorize the use of PCATDs for practical tests or for recency of experience requirements. The studies by Taylor et al. (2001) and Talleur et al. (in press) found that a PCATD and a Frasca FTD were significantly more effective in maintaining recency of experience than a control group that received no training for 6 months. The two groups of pilots who received recency of experience in the two training devices performed at least as well as the group trained in the airplane. This study also showed that 58% of the 106 instrument current pilots in the study failed IPC #1 in an airplane. Thirty-two of these were instrument current then they started their involvement in the study and 56% of these failed an IPC in an airplane. Forty percent of the 15 pilots who were more than 6 but less than 12 months out of currency and who received the recurrent training in a Frasca FTD to regain currency failed an IPC in an airplane. Of the 59 pilots who were more than 12 months out of currency and received about five hours of training in a Frasca and subsequently passed an IPC in a Frasca, 63% failed an IPC in an airplane. The percentage of instrument pilots who failed IPC #1 in the current study, 74%, exceeded the percentage previously observed in Taylor, et al. (2001) and Talleur et al. (in press).

The purpose of the current study is to show the effectiveness and reliability of an FTD, a PCATD, and an airplane in conducting IPCs. To date, 78% of pilots who are legally current have failed the initial IPC. Of the pilots who took IPC #1 in the FTD, 14 of 19 pilots (74%) failed the IPC and of the pilots who took the IPC in the PCATD, 14 of 17 pilots (82%) failed the IPC. The percentage of pilots who failed the initial IPC check flight in the aircraft (78%) was between the percentage for the FTD and the PCATD. The number of subjects in the study who have taken the initial IPC is not sufficient to determine if these results are statistically reliable. The percentage of current subjects failing the IPC in the airplane, 74%, is larger than the percentage of those failing in the Taylor et al. (2001) and Talleur et al. (in press) studies (56%).

Instrument current pilots, regardless of group assignment, are more likely to fail IPC #1 (78%) than to pass it. This finding clearly shows that instrument currency does not necessarily equate proficiency. The data thus far indicates that pilots are more likely to pass IPC #2 in the aircraft than pass IPC #1 in either the PCATD, the FTD, or the airplane. To the extent that all three groups pass rates improve on IPC #2, an overall training effect cannot be ruled out. There is very minimal evidence that pilots retrained to proficiency in the PCATD will pass an IPC #2 in the aircraft, but the data are not sufficient to provide any meaningful statistical inferences at this point (see 2-5 years PCATD row in Table 4).

If a ground-based device is harder to fly than an airplane, then training in such devices may produce a pilot who has an easier time passing an IPC in the aircraft. Current data shows that pilots across all currency groups and experimental groups are as likely to pass IPC #2 as to fail it, regardless of performance on IPC #1. This differs from the results found in the previous project (Taylor, et al., 2001; Talleur et al., in press) where IPC #1 performance was the best predictor of IPC #2 performance.

The data outputs from the FTD and PCATD and the FDRs on board the Sundowner aircraft will be used to examine the possibly different flying characteristics of the different devices and their effects on pilot performance. In addition to the metrics developed and used by Rantanen and Talleur (2001), novel measures based on a time series analysis of the data will be developed. These measurer and analyses will augment the subjective pilot performance evaluation by the CFIIs and help in determining the detailed constituents of pilot performance (or lack thereof) during IPC flights.

ACKNOWLEDGEMENTS

This work is supported under Federal Aviation Administration (FAA) Cooperative Agreement 2001-G-037 and sponsored by the FAA Headquarters, AFS-800. Dennis Beringer serves as the COTR. Views expressed herein do not necessarily represent official FAA positions. We express our appreciation Ms. Mary Wilson who scheduled subjects and to Catherine Trock and Joani Disilvestro for their assistance with the data. We also thank the Institute of Aviation flight instructors who provided instrument training in the FTD, PCATD and the aircraft, the Institute flight instructors who served as IPC check pilots, and the instrument pilots for their participation in the study.

REFERENCES

Lendrum, L., Taylor, H. L., Talleur, D. A., Hulin, C. L., Bradshaw, G. L., & Emanuel, T. W. (2000). *IPC Data Logger Operation Manual.* ARL-00-8/FAA-00-5. Savoy, IL: University of Illinois at Urbana-Champaign

Rantanen, E. M., & Talleur, D. A. (2001). Measurement of pilot performance during instrument flight using flight data recorders. *International Journal of Aviation Research and Development, 1*(2), 89-102.

Talleur, D. A., Taylor, H. L., Emanuel, T. W., Jr., Rantanen, E, M., & Bradshaw, G. L. (in press). Personal Computer Aviation Training Devices: Their Effectiveness for Maintaining Instrument Currency. *The International Journal of Aviation Psychology*.

Taylor, H. L., Talleur, D. A., Bradshaw, G. L., Emanuel, T. W., Rantanen, E. M., Hulin, C. L. & Lendrum, L. (2001). *Effectiveness of Personal Computers to meet recency of experience requirements*. (Technical Report ARL-01-6/FAA-01-1). Savoy, IL. University of Illinois, Aviation Research Laboratory

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

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D. A., Emanuel, T. W., Jr., & Phillips, S .I. (1999) Transfer of training effectiveness of a personal computer aviation training device. *The International Journal of Aviation Psychology, 9* (4), 319-335.

U.S. Department of Transportation. (1997). *Qualification and approval of personal computer-based aviation training devices, AFS-840* (Advisory Circular No: AC 61-126). Washington, DC: Federal Aviation Administration, U.S. Department of Transportation.

U.S. Department of Transportation. (1998, October). *Instrument rating for airplane, helicopter and airship: Practical test standards* (FAA-S-8081-4C). Washington, DC: Federal Aviation Administration, U.S. Department of Transportation.

USE OF TRAINING DEVICES IN GENERAL AVIATION TRAINING PROGRAMS

[#]Michael E. Wiggins and [*]Michael W. Crognale
[#]Embry-Riddle Aeronautical University, Daytona Beach, FL.
[*]University of Nevada Reno, Reno, NV.

While several studies have been done regarding the effectiveness of various training devices in general aviation, not much is known about how they are actually being used by flight schools. This study was designed to gain insight into the way flight schools use training devices. This study surveyed 184 flight schools to gather data about demographics, certification information about their devices, and which tasks are being taught at which level of training in each of the types of devices. Seventy schools responded. The results show that 1) the use of training devices is more prevalent in FAA approved flight schools than other schools, 2) there is some confusion about device certification requirements, 3) training time does not appear to be correlated with the use of these devices, and 4) most of the tasks taught are focused on instrument pilot certification, 5) some schools appear to be using training devices for non-instrument tasks.

INTRODUCTION

Aviation training devices are finding their way into more flight schools than ever before in the past. A recent study of 354 flight schools revealed a total 724 training devices in use (Wiggins, Hampton, Morin, Larssen, & Troncoso, 2002). Of these devices, 381 flight training devices (FTDs), 224 personal computer aviation training devices (PCATDs), and 99 training aids (TA) were reported in use. Most of these devices were used in FAA approved training programs under 14 CFR Part 141 (Part 141) in university-based programs and traditional approved flight schools. Use of these types of devices is not prevalent in schools operating under 14 CFR Part 61 (Part 61). Many of these schools were discovered to be smaller operations where there may not be sufficient resources available to justify the cost or use of these devices.

Increasing capabilities and lowering costs are contributing to this increased use. FTDs and PCATDs have become more usable and realistic, prompting several studies on the usefulness of these devices and how well the training conducted in them transfers to training in airplanes (Lintern, Roscoe, Koonce, & Segal, 1990; Hampton, Monroney, Kirton, & Biers, 1994; Taylor, Lintern, Hulin, Talleur, Emanuel, & Phillips, 1997). These studies all showed positive transfer of training benefits. Studies have been conducting using training devices to teach cognitive skills such as decision-making and situational awareness (Craig, 1999; Wilt, 1997). Benefits from the use of these devices range from the ability to train in less time, train in situations normally considered hazardous in actual flight, to lowering costs.

What is not revealed by any of these studies is how various aviation training devices are actually being used in training programs in general aviation. While guidance exists regarding the certification requirements of these devices (FAA, 1992, FAA 1997), it is not fully known if the devices in use are being used in accordance with these guidelines. Another issue that is not well understood is which areas of operation (AOO) and/or tasks are being targeted for instruction in training devices.

The purpose of this study was to reveal the types of training devices in use, how they are being used to enhance skill and proficiency, which tasks are being taught in these devices, whether or not the devices are appropriate certified and being used in accordance with National Simulator Program (NSP) guidelines, and if they are being used to augment training outside of approved training curricula.

METHOD

This study targeted schools that had previously indicated use of at least one training device in the study by Wiggins, Hampton, Morin, Larssen, & Troncoso (2002). Ultimately 184 schools were targeted for this study. The targeted training curricula were those for private pilot and commercial pilot certification and instrument and multiengine ratings.

A survey was used to collect data in three primary areas: school demographics, device information, and tasks taught in training devices. Part I of the survey collected data regarding school enrollments, hours used by various devices, and training times to certification flown by students. Part II of the survey collected information about the devices, use of these devices in approved training programs, and certification information. Part III investigated which tasks are being taught. In an attempt to standardize terminology, the Practical Test Standards (PTS) were used as the primary reference. Common or similar AOOs from the four PTSs were combined in an attempt to have tasks listed only once. This resulted in 15 AOOs on the survey. Tasks from each PTS were placed under the most appropriate AOO. For each tasks, data was collected on the type of device used, at which level of training that task was taught, and on which learning domain the training was focused (knowledge, skill, or attitude). Representatives from the Federal Aviation Administration from the headquarters office and the NSP office reviewed the instrument. A small pilot test by three chief flight instructors was also conducted. Because the survey was somewhat complex, a set of instructions along with examples of how to complete it correctly were included in the package mailed to each school. The surveys were distributed to the targeted schools along with a cover letter explaining the purpose of the study. Follow up activities consisted of a second mailing and a minimum of three telephone calls to each non-responding school.

RESULTS

Of the 184 schools targeted, 70 (38%) responded: 35 universities, 22 Part 141 schools, and 13 Part 61 schools. Universities had the highest response rate of 53.8% while Part 141 schools and Part 61 schools had response rates of 36.1% and 22.4% respectively. The number of student enrollments totaled 9258 with an average enrollment of 134.2 students per school. Sixty-eight schools provided data about which regulation under which they conduct their training. Forty-eight indicated they conduct training under both Part 61 and 141 while only four conduct training solely under Part 141.Sixteen conduct training solely under Part 61. Table 1 depicts the student training hours to certification.

<u>Table 1</u>

<u>Student Training Hours to Certification</u>

	Private	Commercial	Instrument	Multi-engine
Avg	54.4	104.8	47.0	17.9
Max	75	710	148	87
Min	31	10	12	7
N	52	44	49	41

N= number of schools reporting data

Data were collected regarding how much the devices were used. These data are depicted in Table 2. Averages are for those schools that reported use in each type of device. Data for airplanes is included for reference.

<u>Table 2</u>

<u>Training Hours by Device Type</u>

	Airplanes	FTD	PCATD	TA
Avg/wk/school	442.8	71.1	35.9	51.5
Avg/enroll/school	138.4	165.9	110.4	23.7
Avg/wk/student	3.1	0.4	0.3	2.2
N	65	47	33	6

N = number of schools reporting data on type of devices used.

Data were also collected about use of devices outside of training curriculums for either familiarization or remediation purposes. Eighteen schools reported students who initiate use of training devices on their own for an average of 5.9 hours per student. Fifteen schools reported instructors who initiate use of training devices outside of their curricula for an average of 6.2 hours per student.

In an attempt to see if the use of FTDs by flight schools was significantly correlated with the course completion hours in each of the four courses, some statistical analyses were conducted. The data were divided at the median hour figure and the two groups were compared. The median figure for FTDs hours/week was 10 hours per week with 3 schools reporting 10 hours per week. No significant difference was noted in any of the four courses. The data are shown in Table 3.

Table 3

FTD Use Verses Course Completion Mean Hours

	Priv	Comm	Instrument	Multi
10 or fewer hours/week N = 33	54.1	94.4	46.90	19.7
More than 10hours/wk N = 36	53.5	111.9	45.5	15.1
t-score	.841	1.274	.480	-1.08
Significance	.404	.210	.634	.238

A similar comparison was made based on PCATD use. The median of the hours/week was 1.25 hours/week. Again, a comparison was made between those above and below the median. The data are shown in Table 4.

Table 4

PCATD Use Verses Course Completion Mean Hours

	Priv	Comm	Instrument	Multi
Less than 1.25 hours/week N = 32	55.0	90.1	45.6	15.7
1.25 or more hours per week. N = 32	54.3	121.7	47.5	21.2
t-score	.109	-.878	-.144	-1.371
Significance	.914	.385	.886	.178

As with FTD use, no significant difference was found between group time to completeion for PCATDs. However, these data are correlational and do not address causality. Thus it is possible that subjects completed the course with similar hours because they did use the training devices more and perhaps if they did not, then there might have been a statistical difference between the two groups. An experimental design is required to answer the question of causality.

Questions were asked regarding device certification. The first question asked for the method of certification for a school's FTD. Twenty-six schools reported that their device was approved by a letter of authorization issued after August 1, 1996, 16 indicted their device was approved under the conferred status provision of the guidelines, 7 indicated that they were not sure how their device was certified, and 4 indicated that their device was certified by other means, such as approved in their training course outline or other specific letters of authorization. When asked if they understand the certification requirements and regulations for their FTD, 26 answered they have a complete understanding, 30 answered they have some understanding, and 4 indicated they do not have much understanding. None answered that they have no understanding. When asked if they understand the requirements for continuing use of their FTD, 28 answered that they have a complete understanding, 28 answered that they have some understanding, and 3 answered that they do not have much understanding. No one answered that they had no understanding. When asked if their local Flight Standards District Office (FSDO) was helpful in the approval process for their FTD, 31 answered "very helpful" and 18 answered that they were "somewhat helpful". No one answered that his or her FSDO was not very helpful or were of no help at all.

The data regarding which tasks are taught in each type of device for the four targeted courses is quite lengthy and complex. The data was compiled and displayed in a total of 96 graphs. Each graph depicted the number of students that could have been taught this task. The way in which this number was derived was to add a school's enrollment figure for that course if that school indicated they taught that task in a device. The resulting graphs depicted the number of students by course for each of the tasks listed in the 15 AOOs. A similar method was done to interpret the data about the different learning domains targeted in each device. This data may be more suspect, but does give an indication of the intent of the school. The reason this method of interpreting the data waschosen was to try to offset the fact that some schools have larger enrollments whereas others have only a few students at a time. This seems to give a more meaningful picture than simply the number of schools. Because there is no easy way to condense these graphs for the purposes of this paper, a review of the findings will be given discussing the major findings.

DISCUSSION

With respect to which tasks are being taught in FTDs, the majority seems to be in the area of instrument training. In almost all of the Areas of Operation, instrument students show the highest use in most tasks. This can be expected as most of these devices were designed for instrument training. However, it is interesting to see the number of tasks being taught at the private pilot level. Slow Flight and Stalls is an example of an Area of Operation where private students outnumber students in all other courses. The task Steep Turns, in the Performance Maneuvers Area of Operation is another. In the Ground Reference Maneuvers Area of Operation, there is some indication of use for private pilot training and, to a much lesser degree, in commercial pilot training. Whether or not the increasing number of high quality visual displays that are on newer FTDs, is contributing to this is not known. But it is likely that as newer FTDs with better visual displays are used, training in visual flight maneuvers is likely to increase. This is a potential area for further research, such as is currently ongoing in several places regarding instrument training. FTDs do not appear to be used as much in commercial and multiengine training as they are in private and instrument training, with the exception of those tasks specific to multiengine training.

Looking at the data on KSAs taught in FTDs, there seems to be more emphasis on skills than on knowledge, and very little emphasis on attitudes or decision-making. It is possible that these devices may be unsuitable for attitude or decision-making training or that this area is overlooked or misunderstood by instructors. Since the focus of most training is on the accumulation of knowledge and the development of skills, it may be assumed that decision-making is simply part of those skills and is not looked upon as a separate issue. Airline training in the past decade has evolved to include decision-making and resource management as an integral part of their programs. While it is true that airline training is different from general aviation certification training, it might be worth exploring whether or not some concepts or techniques from airline training can be applied to general aviation.

The use of PCATDs tends to mirror FTD use in most of the Areas of Operation. However, there are some notable exceptions. Takeoffs, Landings, and Go-Arounds is one such Area of Operation. While the total number of students using these devices for this training is rather small, the number of private students is significantly higher than for students training for other ratings or certificates. There is even a small number of students who train the task Rectangular Courses in PCATDs. While this may seem meaningless on the surface, apparently at least one school believes that this training may be of some value. There are even a small number of students who train for multiengine tasks in PCATDs. In the teaching of KSAs in PCATDs, the data show similar trends as with FTD use, with the exception that in some instrument tasks, skills seemed to be emphasized more than knowledge.

Training aids show very little use in most Areas of Operations, with most of that use focusing on knowledge. The data show that some flight schools use these devices, however, so there may be some real value in their use. One factor that may be limiting the use of these devices by schools is the fact that time in such devices cannot be used toward certification. It is not currently known how much students use programs such as Microsoft's Flight Simulator on their own and whether or not this contributes to success in training.

In summary, the data show that use of training devices are mostly in the instrument and private pilot training programs with emphasis on areas that involve airplane systems and procedures, and in instrument flying tasks. Some use is indicated in other tasks but to a much lesser degree. However, the fact that instructors are training students in tasks that are outside tasks related to instrument flying warrants attention and further investigation.

CONCLUSIONS

Many schools, especially those in university-environments and FAA approved schools appear to be using both FTDs and PCATDs a significant amount. Part 61 schools do not seem to use these devices as much. This could because of costs. The data suggests that training devices are used primarily in instrument training, but certainly not limited to that course. The data cannot address the question of whether or not the use of these devices reduces overall flight training time significantly, even though the previously cited research suggests that it can. There appears to be some confusion about training device certification, both for initial certification and continuing use. Most schools felt their FSDO was helpful with the certification of their devices. The data suggests that some schools and/or instructors are experimenting with ways to gain more training value from these devices in courses other than instrument training.

It might be helpful if some simple guidelines for device certification could be developed and distributed to all flight schools. Further controlled experiments are needed to address the question of whether or not flight training hours and thereby costs, can be reduced by the use of FTDs and PCATDs.

REFERENCES

Craig, P. A. (1999). Improving pilot decision-making in situations of high stakes, high stress, and time pressures. (Doctoral dissertation: Tennessee State University, 1998). Dissertations Abstract International, 59.

Federal Aviation Administration. (1997). Qualification and approval of personal computer-based aviation training devices (Advisory Circular No. AC61-126, 5/12/97). Washington, DC: Department of Transportation.

Federal Aviation Administration. (1992). Airplane Flight Training Device Qualification. (AC 120-45A). Washington, DC: Federal Aviation Administration, U.S. Department of Transportation.

Hampton, S., Monroney, W. F., Kirton, T., & Biers, D. W. (1994). The use of personal computer-based training devices in teaching instrument flying: a comparative study. (Final Report). Daytona Beach, FL: Embry-Riddle Aeronautical University.

Lintern, G, Roscoe, S. N., Koonce, J. M., & Segal, L. (1990). Transfer of landing skills in beginning flight training. Human Factors, 32, 319-327.

Taylor, H. L., Lintern, G., Hulin, G. L., Talleur, D. A., Emanuel, & Phillips, S. I. (1997). Effectiveness of personal computers for instrument training. Proceedings of the Ninth International Symposium of Aviation Psychology. Columbus, OH: The Ohio State University.

Wiggins, M. E., Hampton, S., Morin, D., Larssen, A., & Troncoso, A. (2002) A study of training devices used by flight training organizations. Unpublished report prepared for the University Aviation Association under FAA contract No. DTFA 01-01-X-02051. Daytona Beach, FL: Embry-Riddle Aeronautical University.

Wilt, D. F. (1997) The effect of training with authentic activities and instructor scaffolding on the situational awareness of instrument flight students. (Doctoral dissertation, 1996). Dissertation Abstracts International, 58 .

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

An incremental 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, transfer effectiveness ratios (TER) and incremental transfer effectiveness ratios (ITER) were computed for each instrument task and for the time to complete a flight lesson. The data indicate that the PCATD is effective in teaching basic and advanced instrument tasks to private pilots, which replicated the findings of an earlier study by Taylor and colleagues. As a result of prior training in an FTD and a PCATD time to a stage check or an instrument rating flight check flight were less when compared to an airplane Control group.

INTRODUCTION

In an earlier study by Taylor, Lintern, Hulin, Talleur, Emanuel and Phillips (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 criterion performance 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 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. The purpose of the present study is to use an incremental transfer of training research design to measure the effectiveness of a flight training device (FTD) and a Personal Computer Aviation Training Device (PCATD) to determine the point at which additional training in a FTD or a PCATD was no longer effective.

METHOD

Participants

A total of 180 University of Illinois, Institute of Aviation private pilot students, who are enrolled in the Institute's instrument program, will be participating in the study (30 subjects in each group. To date 32 students have completed the study. Each semester the students are assigned equally to the six groups while maintaining a balanced number of subjects across all groups to account for students who drop out of the course prior to completion There are four FTD (Frasca) groups, one PCATD group, and the Control group. All students in AVI 130 and 140 will be involved in the study.

Apparatus

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

Procedure

The instrument training program at the Institute of Aviation is divided into two courses: AVI 130, Basic Instruments and AVI 140, Advanced Instruments. AVI 130 emphasizes aircraft control and instrument departure, enroute and approach procedures, while AVI 140 emphasizes NDB holds and approaches, GPS procedures, and partial panel procedures. The students received 45 hours of lectures during the semester for both courses. For both courses, the students also received 15 flight lessons, each of which were programmed for one lesson per week. Experimental curricula for both courses were developed for the four FTD groups, the PCATD groups and the Control group.

Using an incremental transfer of training design, six groups of subjects were tested in the airplane for proficiency on various instrument flying tasks in both courses. Four of the groups received 5, 10, 15, and 20 hours of prior instrument training in a FTD, respectively. One group received 5 hours of prior training in the PCATD. The prior training was distributed equally between AVI 130 and AVI 140. A Control group received all training in the airplane. Instrument training using the FTD and PCATD was administered to the four FTD groups and the PCATD group during four flight lessons for each semester.

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, Taylor, Lintern, Hulin, Emanuel & Talleur, 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. After all students have completed the study, separate Analyses of Variance (ANOVAs) will be performed to analyze the difference between the six groups on the three dependent measures for both AVI 130 and 140. ANOVAs will be 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, ANOVAs will explore variability in the time to a successful check flight for the AVI 130 and AVI 140 courses as a function of the experimental treatment. To further identify the locus of any significant effects, post–hoc tests will be employed to make specific pair wise comparisons using Tukey's test of significance.

PRELIMINARY RESULTS

At this time, a total of 65 students have completed and taken the final check ride the AVI 130 Basic Instruments course. Table 1 shows the results of the check ride for the six groups for the fall and spring semesters. A total of 41 students passed the check ride on the first attempt and 23 students passed on the second attempt. Six students have been recommended for a remedial course, AVI 102. The total dual flight time to completion is also shown in Table 1. The average course completion time for the Airplane Group is greater the average time for each of the five experimental groups who had prior training in the PCATD or the FTD. The Airplane group required 22.89 hours of dual to complete the course while the five experimental groups required an average of 18.72 hours after prior training in the PCATD or the FTD

A total of 32 students have completed and taken the final check ride (the instrument rating flight check) for the AVI 140 Advanced Instruments course. Table 2 shows the results of the check ride. A total of 18 students passed the check ride on the first attempt and 14 students passed on the second attempt. The 6 students in AVI 140 for the spring semester who were recommended for AVI 102, a remedial course, failed to complete the course during the spring semester and therefore were not given a n instrument rating flight check. The total dual flight time to completion is also shown in Table 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 27.42 hours of dual to complete the course while the five experimental groups required an average of 23.42 hours after prior training in the PCATD or the FTD.

DISCUSSION

The data from the current study thus far indicates that the FTD and the PCATD appear effective in teaching basic and advanced instrument tasks to private pilots. This study systematically replicated 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 and time to the stage check in AVI 130 and to the instrument rating flight check were less for all experimental groups when compared to a Control group trained only in the airplane.

One purpose for conducting an incremental transfer of training study is to determine at what point additional training in the FTD and the PCATD in no longer effective. The amount of data collect thus far does not permit statistical analyses. When additional data are available we hope to be able to answer the question of how can flight schools most effectively use the 10 hours of instrument training time currently permitted by AC No: 61-126 (FAA, 1997). Taylor et al. (1996, 1999) suggested allocating the time to the training of the following instruments tasks: steep turns, intersection holds, ILS, VOR, DME ARC and LOC BC Approaches, NDB holds and approaches, and holds and approaches using partial panel. A study by Taylor, Talleur, Emanuel, Rantanen, Bradshaw and Phillips (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, but the results of the 10 nor the 15 hour groups indicated that it was not an effective use of the additional five hours of time. Flight schools should examine their TCOs to determine where the additional 5 hours could be effectively used. There is also the probability that PCATDs can be used effectively for teaching cross-country procedures where there is the possibility of a one–

to–one transfer of training for time. We are currently investigating the effectiveness of PCATDs for conducting cross-country flights as well as the use of 5 and 10 hours of FTD time to cross-country flight.

ACKNOWLEDGEMENTS

This work is supported under Federal Aviation Administration (FAA) Cooperative Agreement Number 02-G-033 and sponsored by the FAA Headquarters, AFS-800. Dennis Beringer serves as the COTR. Views expressed herein do not necessarily represent official FAA positions. We express our appreciation to Mr. Rick Weinberg for invaluable assistance with flight operations and with student management. Mr. Bill Jones, Mr. David Boyd, Sybil Phillips, and Don Talleur served as check pilots. We also thank the Institute of Aviation the flight instructors and students for their participation in the study.

REFERENCES

Federal Aviation Administration. (1997). Verification and approval of personal computer-based aviation training device (Advisory Circular 61-126). Washington, DC: FAA, Department of Transportation.

Lintern, G., Roscoe, S. N., & Sivier, J. E. (1990). Display principles, control dynamics, and environmental factors in pilot training and transfer. Human Factors, 32, 299-317.

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.

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, Aviation Research Lab.

Table 1
*Flight Lesson 45 Statistics (Fall, 2002 and Spring, 2003)**

	Airplane Only	PCATD 5.00	Frasca 5.00	Frasca 10.00	Frasca 15.00	Frasca 20.00
Number of Students	13	11	9	11	11	10
% First Flight Pass Rate	46.15 (N=6)	72.73 (N=8)	66.67 (N=6)	72.73 (N=8)	81.82 (N=9	40.00 (N=4)
% Second Flight Pass Rate	100.00 (N=7)	100.00 (N=3)	100.00 (N=3)	100.00 (N=3)	50.00 (N=1)	100.00 (N=6)
Students Recommended 102	0	0	1	1	2	2
Total Dual to Completion	22.89 (N=13)	19.40 (N=11)	18.79 (N=9)	19.16 (N=11)	18.74 (N=10)	17.06 (N=10)
Variance Total Dual to Completion	10.68	7.65	5.74	8.71	5.66	11.53

- This lesson is the stage check for AVI 130.

Table 2.
*Lesson 60 Statistics (Spring 2003)**

	Airplane Only	PCATD 5.00	Frasca 5.00	Frasca 10.00	Frasca 15.00	Frasca 20.00
Number of Students	6	6	4	5	5	6
% First Flight Pass Rate	83.33 (N=5)	50.00 (N=3)	100.00 (N=4)	20.00 (N=1)	40.00 (N=2)	50.00 (N=3)
% Second Flight Pass Rate	100.00 (N=1)	100.00 (N=3)	…. (N=0)	100.00 (N=4)	100.00 (N=3)	100.00 (N=3)
Students Recommended 102	1	0	2	1	2	0
Total Dual to Completion	27.42 (N=6)	26.87 (N=6)	25.55 (N=4)	23.28 (N=5)	20.70 (N=5)	20.68 (N=5)
Variance Total Dual to Completion	11.26	5.70	7.10	3.52	4.39	11.45

* Flight Lesson 60 is the Instrument Rating check ride for AVI 140

IMPERFECT AUTOMATION IN AVIATION TRAFFIC ALERTS: A REVIEW OF CONFLICT DETECTION ALGORITHMS AND THEIR IMPLICATIONS FOR HUMAN FACTORS RESEARCH

Lisa C. Thomas, Christopher D. Wickens, and Esa M. Rantanen
Institute of Aviation, Aviation Human Factors Division
University of Illinois, Urbana-Champaign

Automated warning and alert devices such as airborne collision avoidance systems (ACASs) represent a class of automation that is often found to be imperfect. The imperfections can be expressed as the number of false alarms or missed events. Most ACASs are constructed with a bias to prevent misses (which may have catastrophic consequences) and therefore, coupled with a low base-rate of conflict events, create high false alarm rates. In this paper, we review the adequacy of various CDTI warning algorithms that have been proposed and tested in addressing the false alarm issue, and the potential for multiple levels of alerting to mitigate the effects of false alarms on trust and reliance on the CDTI. We suggest new directions for future research, including evaluating the effects of false alarm rates on pilots' use of the CDTI, determining what strategies may enhance pilot tolerance of false alarms, and investigating the use of CDTI in conjunction with air traffic controllers.

INTRODUCTION

Automated warning and alert devices represent classes of automation that are often found to be imperfect (Pritchett, 2001; Stanton, 1994; Sorkin, 1988). The diagnosis of dangerous versus safe conditions is often ambiguous when dealing with uncertain information in a probabilistic world, particularly when the alerting system is forecasting *future* situations in uncertain environments. Such circumstances characterize airborne collision avoidance systems (ACASs) such as the traffic information and collision avoidance system (TCAS), which is in operational use today, or longer range planning systems, such as the Cockpit Display of Traffic Information (CDTI), which is still under development (Johnson, Battiste, & Bochow, 1999; Battiste & Johnson, 2002; Johnson, Jordan, Liao, & Granada, 2003).

In a general sense, the imperfection of any warning system can be expressed in signal detection terms as errors of misses (a true dangerous situation is not detected) or false alarms (a safe situation triggers a warning). Misses and false alarms (FAs) trade off against each other; an extremely sensitive system that almost never misses a potential conflict necessarily has a high false alarm rate (FAR). Because misses have potentially catastrophic consequences to aircraft crews and passengers as well as often negative legal implications to systems manufacturers, most warning systems are constructed with a bias to prevent misses, consequently increasing the FAR (Parasuraman, Hancock, & Olofinboba, 1997; Kuchar, 2001). The FAR can be quite high if the base rate of events to be detected is low (Krois, 1999). However, high FAR has significant negative repercussions too, and may lead to operator mistrust and consequent "disuse" of automation (e.g., Sorkin, 1988; Parasuraman & Riley, 1997). The effect of FAs on human performance is therefore the primary human factors issue associated with automated alerting systems.

However, relatively little research appears to have examined the relative consequences of FAs *versus* misses in influencing human trust and reliance on automated alerts. A few recent studies in the context of automobile warnings suggest that FAs may indeed be more degrading of trust than misses (Gupta, Bisantz, & Singh, 2001; Cotté, Meyer, & Coughlin, 2001). When an alarm is annunciated and directs the attention of the operator away from other tasks, and this alarm turns out to be false, the operator has wasted time and effort in dealing with it and is more likely to lose trust in a system that demands this extra effort. A miss, on the other hand, is by definition not annunciated and therefore the operator has spent no energy in dealing with it and is not likely to even know that a real event exists and was missed. Unless the operator is somehow prompted to determine whether the system missed some critical events, the operator is likely to maintain his/her initial level of trust in the system. It must be noted here that this discussion pertains only to human performance, that is, trust; although misses that remain unknown to the user do not erode trust, they are hardly desirable from the system performance perspective.

THE PROBLEM DEFINED

It should also be noted that the issue in conflict detection algorithms is not so much misses per se as it is delayed issuance of alarms. A system that detects conflicts based on continuously updated information about the location and trajectory of surrounding aircraft will always detect a conflict eventually. If the conflict actually exists, the evidence for it will eventually cross the critical threshold for an alert. We therefore define a miss by the conflict detection system as an alert that is produced at such a late time that the pilot has to take immediate action (if any action can be taken at all) to resolve the conflict.

Figure 1 provides a schematic illustration of the issues in selecting alarm thresholds for a CDTI, plotting the separation between two aircraft as a function of the passage of time during a potential conflict episode. Time 0 is some arbitrary time prior to the point of closest passage between the two aircraft, defined as the look-ahead time (LAT). The solid line shows the nominal prediction, illustrating the steadily decreasing distance to closest passage (DCP), followed by the increase thereafter. The instant any trajectory crosses the minimum threshold of 3 (or 5) miles of separation, (or any other arbitrary separation distance) a formal *conflict* is defined. In the CDTI, the pilot should be alerted with a sufficient margin of time before conflict occurs so that she or he is able to non-aggressively maneuver in any of the three axes of flight to avoid it.

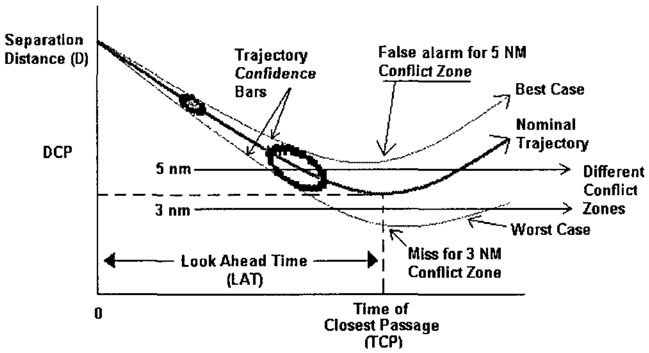

Figure 1. Representation of the evolving space and time aspects of a conflict.

The nominal trajectory represents the expected evolution if neither aircraft alters its speed or heading from that observed at time 0. However, such deterministic behavior is rarely observed. The two "eggs" in Figure 1 represent the anticipated variability in both speed and lateral position around the nominal trajectory (Magill, 1997). This variability increases with increasing time. These "eggs" can be thought of as confidence intervals (e.g., 90%). The lighter lines surrounding the nominal trajectory represent the confidence intervals on lateral separation, in which the "best case" line is the maximum predicted separation distance at closest passage, and the "worst case" line is the minimum predicted distance at closest passage. The growth of uncertainty over time represents the impact of winds or other factors that cannot be predicted with certainty.

Now consider a warning that might be given at time 0, defining an LAT to closest passage or to another event, such as a loss of separation. If, for example, the warning is based on the nominal trajectory for a 5 mile protected zone, and then a "best case" trajectory actually occurs, this would lead to a false alert. On the other hand, if the protected zone is 3 miles, no warning will be given if it is based on the same projected nominal trajectory, and if a "worst case" trajectory actually occurs the system has produced a miss (or at best, a delayed alarm).

The designer must decide whether to issue the warning based upon the nominal trajectory, or some worst case value (90%, 95%, etc.), by balancing the costs of delayed alerts ("misses") versus the costs of false alerts (Yang & Kuchar, 1997). Complicating the design issue further is the LAT. If the trajectory is deterministic, then any LAT will produce equal (and perfect) accuracy. Furthermore, if LAT is very short, accuracy can also be nearly perfect. However the growth of uncertainty with longer LATs, shown by the increasing range of confidence intervals in Figure 1, implies that the longer the LAT, the greater the tradeoff between late alerts and false alerts. Yet, as noted above, the LAT must be great enough to allow the pilot sufficient time to maneuver in a non-aggressive fashion.

As Figure 2 illustrates, the LATs can be categorized into three basic categories according to proposed use of the conflict detection system: Emergency, which generally requires immediate and often constrained actions (e.g., vertical maneuvers only) to resolve the detected conflict; Tactical, which allows the pilot enough time to consider several resolution options and then choose one to implement; and Strategic, which provides a significantly larger amount of time to create very slight modifications of the flight plan in order to avoid conflict with the least impact to the existing flight plan. TCAS' Resolution Advisories operate within Emergency LATs, which are expected to produce the highest hit rate but may still suffer the effects of FAs. The CDTI developed at UIUC uses an algorithm that provides 45 seconds of warning before loss of separation (see Alexander & Wickens, 2002). In both cases, the pilots are expected to take immediate action (usually a time-efficient vertical maneuver) to resolve the imminent conflict.

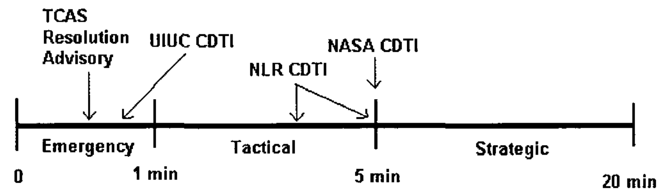

Figure 2. Representation of Look-Ahead Times

NLR and NASA have created CDTIs using algorithms that provide 3 to 5 minute LATs (see Hoekstra & Bussink, 2003; Johnson, Battiste, & Bochow, 1999). Pilots are alerted to a detected conflict, but have several minutes to determine the best course of action to resolve the conflict with minimal impact to flight characteristics such as the time schedule, fuel costs, and physical maneuvers available. When the pilots have more time to create conflict resolution plans, they can utilize maneuvers in any of the three flight dimensions (vertical, lateral, and airspeed), which in turn allows them to create more efficient (albeit more complex) resolutions. With a 3-5 minute LAT, however, the system is subject to both misses and false alarms depending on how accurately the algorithm predicts the trajectory. Algorithms with longer LATs have been evaluated (see Magill, 1997) for strategic flight planning use, but it is likely that with the increase in uncertainty at such long LATs the rate of both false alarms and misses will be prohibitively high and will not produce a useful tool when it comes to planning for projected conflicts.

Thus, as is evident from the analysis, the joint influence of the three parameters (the protected zone size, the LAT, and the assumptions about the growth of uncertainty with time) will affect the sensitivity of discriminating predicted conflicts from non-conflicts, and hence the extent of the tradeoffs between the two negative events of false alerts and misses or late alerts. We will review some empirical work pertaining to these factors and their human performance implications next.

LIMITATIONS OF CURRENT RESEARCH

In the process of gathering information on proposed conflict detection algorithms, we reviewed over 40 articles which contained one or more of the following: (1) a description of an algorithm, (2) analytical validation of an algorithm, or (3) validation of an algorithm by pilot-in-the-loop (PIL) simulations. This review revealed that very few of the algorithms have been validated in realistic free flight simulations with PIL performance data. For the purposes of this paper, we have chosen to illustrate six PIL studies (Table 1), which are representative of the type of studies that have been conducted on the different algorithms mentioned above, along with a breakdown of key characteristics of the studies.

The NASA studies that appear in the first three rows in Table 1 show the range of approaches in implementing and evaluating a CDTI containing a single conflict detection algorithm (Yang & Kuchar's 1997 algorithm), which detects conflicts for 5 NM protected zones. All of these studies used multi-level alerts and reported PIL performance data, but only one (Johnson et al., 1997) varied uncertainty growth parameters and none considered false alarms.

The fourth study (Wing, Barmore, & Krishnamurthy, 2002; see also Wing et al, 2001) is an investigation of a CDTI that incorporates features of two probabilistic algorithms (Yang & Kuchar's 1997 algorithm & NLR algorithm) to detect conflicts using different sources of information (state or intent), while the fifth study (Hoekstra & Bussink, 2003) implemented the NLR algorithm alone. Neither of these studies specified any uncertainty parameters nor manipulated FAR as an independent variable.

The final set of studies (from the University of Illinois) used a non-probabilistic algorithm developed at the University. These experiments are the only ones discovered that manipulated the protected zone and lateral uncertainty as independent variables. In addition, only Wickens, Gempler, and Morphew (2000) involved misses as an experimental variable.

The reviewed research shows that pilots can use ACAS technology to successfully aid in-flight separation and also that pilots report high subjective approval ratings of the availability of CDTI information. However, there are several major areas of research that have not yet been addressed by simulations of CDTI and conflict detection algorithms. The simulation-based validations reviewed here tended to be limited in scope with respect to consideration of a variety of conflict situations, alerting and traffic display characteristics, and conflict detection capabilities. Furthermore, the sample size has been generally small, potentially resulting in lack of statistical power in making strong general conclusions. While there has been some discussion of FA and delayed alarm rates (Yang & Kuchar, 1997; Kuchar, 2001; Hoekstra & Bussink, 2003), we have found only one study that has addressed "missed" conflicts (or delayed alerts) as a variable (Wickens et al., 2000), and none that have investigated FA effects on pilot preference and trust directly, much less manipulated FAR (as dictated by alarm threshold or LAT) as variables in a study.

Our analysis of the larger set of algorithm studies (from which Table 1 is derived) reveals that the three most critical variables for affecting the balance of FAs versus late alarms (depicted in Figure 1) are (1) LAT, as a longer LAT produces more FAs, (2) the size of the minimum separation boundary, where the larger the boundary, the more FAs produced, and (3) the assumptions that are made about the growth of uncertainty (see also Magill, 1997). Yet Table 1 reveals little consistency across these variables between studies (see in particular the "Uncertainty Growth Parameters" column), and no systematic manipulation of them in the PIL studies.

Some prior research has suggested that a key feature for mitigating the negative consequences of false or "nuisance" alarms is the capability of providing graded levels of alerting, such that the user would be less distressed if an alert at the lowest level of predicted danger proves to be incorrect (Sorkin, Kantowitz, & Kantowitz, 1988; St. John & Manes, 2002). As shown in column 5, the six studies described in Table 1 used multiple levels of alerting (between 2 and 5 alerting levels) to indicate the relative urgency of the alarm. However, none of these studies directly compared different numbers of alerting levels to each other within a single study. In sum, we have found no consistency in the implementation of the multiple level alerts across studies, and have found no studies that have investigated the *optimal* number of alert levels.

FUTURE RESEARCH ISSUES

Based on our review of ACAS literature, we will make several recommendations for future research. First, since it probably is not possible to determine a fixed threshold for an "acceptable" FAR due to the complexity of constructs such as trust and workload and the innumerable factors affecting them (see Parasuraman & Riley, 1997), as well as the diversity of the operational environments and settings in which alerting systems are used, research focus should be on the operators' tolerance for the inevitably high FAR and the role of training and system design in improving that tolerance. The FA tolerance could be increased by improving pilots' general awareness of the traffic situation on one hand, and the accuracy of their mental model of the algorithms of the collision alert system on the other. Second, since unaided humans are notoriously bad at estimating probabilities and making judgments based on probabilistic information (Kahneman & Tversky, 1984), the operators' performance could be improved by displaying probabilistic information to them in a form that is easy to perceive and understand and that can be readily used in their tasks, such as in the form of graded levels of alerting. Finally, the role of CDTI in the free flight environment will be drastically different from that of TCAS. It is hence crucial to examine its use in concurrence with ATC procedures and controllers' tasks. The congruence of planning and conflict detection algorithms of CDTI and ATC automation tools will have a substantial impact on the performance of both pilots and controllers.

SUMMARY AND CONCLUSION

In this paper, we have considered the adequacy of various CDTI warning algorithms that have been proposed and tested in addressing the FA issue. We also noted the important distinction between testing the algorithm (software) itself, and testing the pilot's use of the algorithm in a conflict avoidance PIL simulation or in operational context. Finally, we described a framework for addressing the FA issue from the perspective of the pilot's decision-making when interacting with CDTIs.

It is apparent that the present research findings on the effects of FAs on human trust, workload, and performance in conjunction with ACAS technology must be considered in the light of the operational environment in which the systems are to be used. For example, the envisioned use of CDTI as a strategic planning tool with relatively long LAT will likely result in very different pilot responses to FAs than what has been found in immediate conflict avoidance settings. Such complex environments, however, place substantial demands to the design of experiments, which must manipulate or control all the relevant independent variables and accurately measure the dependent variables. In the latter category, what ultimately matters most is human performance, posing further challenges to the characterization and measurement of apposite parameters.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the Federal Aviation Administration (no. 02-G-032) to the third author. The FAA technical monitor for this grant is Dr. William K. Krebs, AAR-100. The views expressed in this paper are those of the authors and do not necessarily reflect those of the FAA

REFERENCES

Alexander, A., & Wickens, C. D. (2001) *Cockpit display of traffic information: The effects of traffic load, dimensionality, and vertical profile orientation* (ARL-01-17/NASA-01-2), UIUC, Aviation Research Lab.

Battiste, V., & Johnson, N. (2002) An operational evaluation of ADS-B and CDTI during airport surface and final approach operations. *Proc. 46th Annual HFES Meeting.*

Cotté, N., Meyer, J., & Coughlin, J. F. (2001). Older and younger drivers reliance on collision warning systems. *Proc. 45th Annual HFES Meeting.*

DiMeo, K, Sollenberger, R., Kopardekar, P., Lozito, S., Mackintosh, M-A., Cardosi, K., & McCloy, T. (2002) *Air-ground integration experiment* (DOT/FAA/CT-TN02/06).

Gupta, N., Bisantz, A. M., & Singh, T. (2001). Investigation of factors affecting driver performance using adverse condition warning systems. *Proc. 45th Annual HFES Meeting.*

Hoekstra, J., & Bussink, F. (2003) *Free flight: How low can you go?* DASC Report.

Johnson, W. W., Battiste, V., Delzell, S., Holland, S., Belcher, S., & Jordan, K. (1997) Development and demonstration of a prototype free flight cockpit display of traffic information. *Proc. 1997 SAE/AIAA World Aviation Congress.*

Johnson, W. W., Battiste, V., & Bochow, S. (1999) A cockpit display designed to enable limited flight deck separation responsibility. *Proc. 1999 World Aviation Conference.*

Johnson, W. W., Jordan, K., Liao, M., & Granada, S. (2003) Sensitivity and bias in searches of cockpit display of traffic information utilizing highlighting/lowlighting. *Proc. 12th International Symposium on Aviation Psychology.*

Kahneman, D., & Tversky, A. (1984) Choices, values, and frames. *American Psychologist, 39*, 341-350.

Krois, P. (1999). *Alerting systems and how to address the lack of base rate information.* Unpublished manuscript.

Kuchar, J. K. (2001) Managing uncertainty in decision-aiding and alerting system design. *Proc. 6th CNS/ATM Conference.* Taipei, Taiwan.

Magill, S. A. N. (1997). Trajectory predictability and frequency of conflict avoiding action. *Proc. CEAS Free Flight Conference.* DERA.

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

Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors. 39*(2), 230-253.

Pritchett, A. (2001) Reviewing the role of cockpit alerting systems. *Human Factors and Aerospace Safety, 1*(1), 5-38.

St. John, M., & Manes, D.I. (2002) Making unreliable automation useful. *Proc. 46th Annual HFES Meeting.*

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

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

Stanton, N. (1994). *The human factors of alarm design.* Bristol, PA: Taylor and Francis.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors Journal, 2*(2), 99-126.

Wickens, C. D., Xu, X., Helleberg, J, & Marsh, R. (2001) Pilot visual workload and task management in free flight: A model of visual scanning. *Proc. 11th International Symposium on Aviation Psychology.*

Wing, Adams, Duley, Legan, Barmore, & Moses, (2001) "Airborne use of traffic intent information in a distributed air-ground traffic management concept: Experiment design and preliminary results." NASA/TM-2001-211254

Wing, D. J., Barmore, B. E, & Krishnamurthy, K. (2002). *Use of traffic intent information by autonomous aircraft in constrained operations.* AIAA.

Yang, L. C. & Kuchar, J. K. (1997). Prototype conflict alerting system for free flight. *Journal of Guidance, Control, and Dynamics, 20*(4), 768-773.

Table 1. *Summary of six pilot-in-the-loop studies that incorporated different conflict detection algorithms into CDTIs for use in free flig* *imulations.*

Reference	Algorithm type	Attributes varied?	Uncertainty growth parameters	Multi-level alerts	t-in-loop imary	False Alert discussion?
Johnson, Battiste, & Bochow (1999)	Yang & Kuchar algorithm used in the CDTI	Protected Zone (PZ) = 5 NM, +/-1000 ft; intent info was either shared or unshared; 2 *types of conflicts = true or* "apparent"; initial altitudes were either separated or not separated	No error factors	3 levels: 1. Moderate threat in the future, or low immediate threat 2. *Moderate immediate threat*, high threat in the future 3. High near-future threat	ews flew either or low flight-deck onsibility *narios; positive* ect responses to lability of info, rs, display options	
Johnson, Battiste, Delzell, Holland, Belcher, & Jordan (1997)	Yang & Kuchar algorithm used in the CDTI	PZ = 5 NM, +/- 1000 ft vertical	Along-track error standard deviation (s.d.) of 15 knots, cross-track error s.d. of 1 NM; probabilities of turns or altitude changes included (4 turns/hr, 4 alt changes/hr)	5 levels: 1. Low: no alert/info only 2. Moderate: no alert info only 3. High: alert 4. Very high + Traffic advisory 5. Very high + Resolution advisory	rews flew either anced (included lution tools) or c CDTI condition; ew 8 scenarios w/ ast one conflict 1	Authors state explicitly that no false alarms were programmed into the scenarios
DiMeo, Sollenberger, Kopardekar, Lozito, Mackintosh, Cardosi, & McCloy (2002)	used Yang & Kuchar (1997) logic, overlaid on TCAS	PZ = 5 NM, +/- 1000 or +/- 2000 ft as appropriate (if ownship alt. > 29,500 ft, vertical separation is +/- 2000 ft)	Uncertainty parameters not specified (assumed to be same as described in Yang & Kuchar, 1997)	4 stages: • First two levels are CDTI-based • Last two levels are TCAS Traffic Advisory and Resolution Advisory	r sets of /copilot flew 4 rent conditions, 3 s each	
Wing, Barmore, & Krishnamurthy (2002); Wing, Adams, Duley, Legan, Barmore, & Moses (2001)	Combination of Yang & Kuchar logic (state + intent) plus NLR/Hoekstra logic (state only)	PZ = 5 NM, +/-1000 ft; LAT = 5 minutes for state info, 8 minutes for intent info; 3 categories of conflicts = state-only, intent-only, and "blunder"	Uncertainty parameters not specified (assumed to be same as described in Yang & Kuchar, 1997)	3 alerting stages: 1. Inform pilot of potential conflict, no action required 2. Conflict detected, action by OS required 3. LOS has occurred, immediate action by OS required	ommercial pilots 3 en-route trials 1 of 4 scenarios ical vs. strategic ration mode, low high ops plexity)	
Hoekstra & Bussink (2003)	NLR state-based conflict detection algorithm	PZ = +/- 1000 ft, 5 NM en-route, 3 NM near terminal	Uncertainty parameters not specified - traffic generated by NLR software "Traffic Manager"	2 levels of urgency: 1. LOS occurs between 5-3 min before closest point of approach 2. LOS occurs less than 3 min to closest point	it sets of pilots en-route descent terminal roach trials with vs. high traffic, managed vs. free t terminal roach	False alarms are assumed to be minimized by setting LAT to 5 min (see Magill, 1997)
Wickens, Gempler, & Morphew (2002); Wickens, Xu, Helleberg, & Marsh (2001); Alexander & Wickens (2001)	Geometric linear extrapolation of vertical and curvilinear extrapolation of lateral (rate of turn).	PZ = lateral and vertical dimensions varied between experiments.	Wickens, Gempler, and Morphew introduced lateral uncertainty. Not quantitatively varied.	Two levels: 1. Predicted loss of separation within 45 seconds 2. Actual loss of separation.	s found useful predictor of e trajectory, and lict specific mation about t of closest sage.	Wickens, Gempler, & Morphew produced conflict misses; the only such study to do so.

Appendix II

Human Factors General Aviation

Research Requirements

Below are the requirements that pertain to the projects listed in Appendix I

Research Requirement

Credit for Instrument Rating in a FTD

Developing and Validating Criteria for Constraining False & Nuisance Alerts for Cockpit Display of Traffic Information Avionics

FAA/Industry Training Standards (FITS)

Human Error and General Aviation Accidents: A Comprehensive, Fine-grained Analysis Using HFACS

Below are the requirements for two new start FY04 projects

Low Visibility and Visual Detection

Primary Flight Displays, Terrain, Overlays/layers, Perspective Displays

Requirement ID: 767

Sponsor Organization: AFS POC: Anne Graham

Requirement Title: Credit for Instrument Rating in a Flight Training Device

Funded Requirement:
- FY02: Yes
- FY03: Yes
- FY04: Yes
- FY05: Yes

Requirement Statement: This information is required for the revision of FAR 61-141, specifying the credit hours for which various Flight Training Devices (FTDs) may be used in lieu of actual flight.175

Background: 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 posible 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 inthe 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: a. Evaluate all seven levels of FTDs, recategorizing them as necessary by shared characteristics (i.e., fidelity fo physical/visual/flight replication) b. Develop a system for measuring and recording a range of pilot performance within the areas of aircraft handling, navigation, and emergency procedures. c. Measure the performance levels of students from each of the seven FTD categories. d. Determine the pont at which performance levels in an aircraft meet pilot certification standards???

Output: Final report that provides guidance as to what specific maneuvers (initial private and initial instrument training) can be completed in the FTD and/or PCATD in lieu of flight time. Provide guidance as to whether FTD and/or PCATD can be used for recurrent training and instrument proficiency checks in lieu of flight time.

Regulatory Link: none

Requirement ID: 860

Sponsor Organization: AIR POC: Colleen Donovan

Requirement Title: Developing And Validating Criteria for Constraining False & Nuisance Alerts For Cockpit Display Of Traffic Information Avionics

Funded Requirement:

- FY02: Yes
- FY03: Yes
- FY04: No
- FY05: No

Requirement Statement: The objective of this project is to develop and validate criteria for constraining false and nuisance alerts for cockpit displays of traffic information (CDTI), based on what is known about other alerting algorithms and human performance issues with alerting systems, trust, situation awareness and workload. Where objective criteria are not possible, subjective means may be recommended provided they are established to be reliable and valid measures. These criteria are to be included as minimum requirements in the RTCA Minumum Operational Performance Standards document or an FAA Technical Standard Order for CDTI. Both of these documents are used by avionics manufacturers to develop their systems, and FAA aircraft certification specialists who evaluate the systems. The project should be focused on developing these objective and subjective measures as minimum certification criteria, based on research and data, for approving the Free Flight technologies known as Cockpit Displays of Traffic Information (CDTI). The CDTIs may be either stand-alone units or as part of an integrated ADS-B CDTI/Traffic Collision Avoidance System (TCAS).1148

Background: It can be argued that the efforts to modernize the NAS and enhance both capacity and safety of the nation's air transportation system are presently technology-driven and that human factors contributions to these efforts have fallen behind the demand. The reason for this situation is apparent: The task environments in which the personnel ultimately responsible for the safe and efficient functioning of the NAS (i.e., pilots, airline dispatchers, air traffic controllers and –managers) work have increased in complexity with increase in automation applications. Consequently, scientific investigation of the impact of new technologies has become increasingly difficult due to the escalating number of variables and their interactions in the present operational environments and the shift from overt performance (i.e., manual control) to predominantly covert behavior (i.e., supervisory control) of the operators. Several constructs that attempt to describe the complex and mostly covert behaviors have been introduced. The most significant of these is situation awareness (SA), but trust and workload associated with automation are of concern as well. The measurement of these constructs is problematic, yet of critical importance. May want to insert something here talking about the numerous problems with alerting systems and false alerts- impact on human performance- pilots turn them off, ignore them (boy who cried wolf) etc. This research will span a period of three years, with three distinct phases. Each phase may be considered individually for support, but the latter phases will depend on successful completion of the previous phases. Phase 1 and the first year efforts will focus on data gathering and understand how similar issues were solved with other flight deck alerting systems, such as TCAS, enhanced ground proximity warning systems (EGPWS) and wind shear alerts. This phase will include exhaustive review of the certification standards, requirements and guidelines related to false alerts and alerting criteria published in RTCA MOPS and TSOs for the systems mentioned above. The background and basis for the currently published standards should also be examined, as well as research literature pertaining to human performance issues with alerting systems associated with situation awareness, trust, and workload. The interactions of these constructs will also be examined, with an objective of identifying common underlying structures or mechanisms. This will include a review and evaluation of the Aviation Safety Reporting (ASRS) literature associated with TCAS problems, as well as other TCAS issues in order to uncover lessons learned. Special emphasis will be paid to the three "key references" listed at the end of the paper, as a potential means to develop certification standards to enable the evaluation of traffic collision alerting systems (e.g., CDTI ADS-B, TIS, and TCAS). These key reference papers propose the use of Signal Detection Theory (SDT) methodology as a means to evaluate alerting systems and separate the impact of various decision biases. SDT can be used to study the impact of changes to the decision threshold, and also the impact of changes to the a priori base rate events in the real world. The authors of these key references establish the importance not only of high hit rates and low false alarm rates, but also of the importance of high posterior probabilities of a true alarm. Additionally, they also propose a means to access the impact of these changes, despite the fact that only a handful of airplanes are equipped with ADS-B/CDTI systems, and thus it is difficult to determine the base rate information for these events, which is required to determine the posterior probabilities. Thus, one path of pursuit towards objective criteria to evaluating the CDTI alerting system is by attempting to apply the methodologies proposed and developing recommended certification criteria for the alerting systems hit rates, false alarm rates, and posterior probabilities. This methodology may prove effective in developing objective criteria for evaluating the appropriateness of an alerting system on the "trust/use/misuse/abuse" dimension. Additional methodologies and criteria would need to be developed to evaluate the situation awareness and workload dimensions.

Output: Year 1 1. Documentation review: a) empirical human factors results relevant to alerting systems, available in the public domain (journal articles, conference proceedings, and government reports); b) certification standards, requirements and guidelines related to false alerts and alerting criteria published in RTCA MOPS and TSOs for cockpit alerting systems; c) comparison of the alerting algorithms of TCAS, CDTI, CA, and URET and examination of their congruence with pilots' and controllers' tasks and mental models; d) previous ASRS analyses on alerting system related incidents to determine if yet another ASRS analysis is warranted; e) literature on human factors certification for guidelines for development of certification criteria for CDTIs; f) identification of other data sources (e.g., from demonstrations and simulations or from operational environments) that would allow for further examination of relevant human factors issues outside of a laboratory. 2) Examination of the roles of cockpit alerting systems. This subtask will examine the roles of a number of automatic alerting systems (GPWS, TCAS, wind shear alert) and the impact of these on the respective certification criteria of the alerting systems. 3) Development of measures and criteria for collision avoidance system evaluation. This subtask involves a comprehensive evaluation of available measures of machine, human, and human-machine system performance as they pertain to collision avoidance systems, identification of primary and secondary measures, and evaluation of empirical support for the latter. We will also examine possible sources and justification for criteria for the measures. 4) Develop designs and protocols for experiments. Based on findings from the literature review, we will develop experimental designs and protocols aimed at investigation of the most critical issues relevant to human factors certification of CDTIs and to address possible controversies in the alerting system literature. 5) Conduct Experiment 1. The goal of this component of the project is be to develop a cognitive model of the features of unaided conflict prediction, that is, pilot prediction made without the aid of intelligent automation.

Regulatory Link: none

Requirement ID: 887

Sponsor Organization: AFS POC: Bob Wright

Requirement Title: FAA/Industry Training Standards (FITS)

Funded Requirement:
- FY02: No
- FY03: No
- FY04:
- FY05:

Requirement Statement: 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 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

Background: Within the past five years, avionics manufacturers have developed a large number of general aviation products to improve pilots'' situational awareness. Although these products are advertised to enhance safety and efficiency, there are a number of skeptics who question the utility of these products. In fact, many in the general aviation community believe that some of these aviation products are training intensive and present complex human factors issues that must be resolved to obtain the full safety benefits or, in some cases, to avoid creating new safety issues. The purpose of the FAA/Industry Training Standards (FITS) program will be to develop a flexible but robust general aviation training programs that can be tailored to integrate different technologies into any aircraft platform.

The FITS training program would be web-based documentation repository that would contain the FAA/Industry training standards most up-to-date information to support general aviation guidelines, standards and certification, and other materials. The FITS database would contain training standards for specific technologies by aircraft type. For example, a flight instructor preparing an instrument student would access the FITS website and select the instrument training module standard that matches the aircraft type and avionics installed in the aircraft. The FITS instrument-training program would contain real-world scenarios based on problem solving and case study examples with defined metrics for evaluation on aeronautical decision making, information management and risk management.

Output: Near term products:
Establish web site that will distribute FITS information, Establish template for FITS products, Publish Advisory Circular on FITS, Aviation safety inspector training and guidance, Designated examiner guidance.
Future Products:
Transition training, Type specific aircraft training, Type rating training, Special training (i.e. R-22, MU-2), Recurrent training, Currency requirements, Equipment training (i.e. GPS, HITS, MFD/PFD), Specific avionics equipment training, Ab-initio training for professional pilots, Ab-initio training for non-professional (enthusiast) pilots, First officer training, Designated examiner/FAA inspector training, Flight instructor renewal, Possible 14 CFR part 135 training

Regulatory Link: none

Requirement ID: 868

Sponsor Organization: AFS POC: Robert Wright

Requirement Title: Human error and general aviation accidents: A comprehensive, fine-grained analysis using HFACS

Funded Requirement:

- FY02: previously “Causal factors of accidents and incident attributed …”
- FY03: Yes
- FY04: Yes
- FY05: No

Requirement Statement: 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

Background: Humans by their vary nature make mistakes; therefore it is unreasonable to expect error-free human performance. It is no surprise then, that human error has been implicated in a variety of occupational accidents, including 70% to 80% of those in civil and military aviation (O'Hare, Wiggins, Batt, & Morrison, 1994; Yacavone, 1993). In fact, while the number of aviation accidents attributable solely to mechanical failure have decreased markedly over the past 40 years, those attributable at least in part to human error have declined at a much slower rate (Shappell & Wiegmann, 1996). It appears that interventions aimed at reducing the occurrence or consequences of human error have not been as effective as those directed at mechanical failures. Clearly, if accidents are to be reduced further, more emphasis has to be placed on the genesis of human error as it relates to accident causation.

The predominant means of investigating the causal role of human error in aviation accidents remains the analysis of accident and incident data (Shappell & Wiegmann, 1997). Unfortunately, most accident reporting systems are not designed around any theoretical framework of human error. Indeed, most accident reporting systems are designed and employed by engineers and front-line operators with limited backgrounds in human factors. As a result, these systems have been effective at identifying engineering and mechanical failures, whereas the human factors component of these systems are generally narrow in scope. Furthermore, even when human factors are specifically addressed, the terms and variables used are generally ill defined and the data structures poorly organized. Postaccident databases are therefore not conducive to a traditional human error analysis, making the identification of intervention strategies onerous (Wiegmann & Shappell, 1997).

What is required therefore, is a general human error framework around which new investigative methods can be designed and existing postaccident databases restructured. However, previous attempts to apply error frameworks to accident analysis have met with encouraging, yet limited, success (O'Hare et. al., 1994; Wiegmann & Shappell, 1997). This is due primarily to the fact that performance failures are influenced by a variety of human factors that usually are not addressed by traditional frameworks. With few exceptions (e.g., Ramussen, 1982), human error taxonomies do not consider the potential adverse mental and physiological condition of the individual (e.g., fatigue, illness, attitudes, etc.) when describing errors in the cockpit. Furthermore, latent errors committed by officials within the management hierarchy, such as line managers and supervisors are often not addressed, even though it is known that these factors directly influence the condition and decisions of pilots (Reason, 1990). Therefore, if a comprehensive analysis of human error is to be conducted, a taxonomy that takes into account these multiple causes of human failure must be offered.

A comprehensive Human Factors Analysis and Classification System (HFACS) has recently been developed to meet these needs (see Figure 1). This system, which is based upon Reason’s (1990) model of latent and active failures addresses human error at each of four levels of failure: 1) unsafe acts of operators (e.g., aircrew), 2) preconditions for unsafe acts, 3) unsafe supervision, and 4) organizational influences. The HFACS framework was originally developed for the U.S. Navy and Marine Corps as an accident investigation and data analysis tool. Since its original development however, HFACS has been employed by other military organizations (e.g., U.S. Army, Air Force, and Canadian Defense Force) as an adjunct to preexisting accident investigation and analysis systems. To date, the HFACS framework has been applied to over 1,000 military aviation accidents yielding objective, data-driven intervention strategies while enhancing both the quantity and quality of human factors information gathered during accident investigations (Shappell & Wiegmann, 2001).

Other organizations such as the FAA and NASA have explored the use of HFACS as a complement to preexisting systems within civil aviation in an attempt to capitalize on gains realized by the military. These initial attempts, performed both at the University of Illinois and the Civil Aerospace Medical Institute (CAMI) have been highly successful and have shown that HFACS can be reliably used to analyze the underlying human factors causes of both commercial and general aviation accidents (Shappell & Wiegmann, 2001; Wiegmann & Shappell, in press). 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. Indeed, AFS-800, the Aeronautical Decision Making (ADM) JSAT and the General Aviation Data Improvement Team (GADIT) have acknowledged the added value and insights gleaned from these HFACS.

To date, however, these initial analyses using HFACS have generally been performed at a global level and several questions remain concerning the underlying nature and prevalence of different error types. In fact, AFS-800, the ADM JSAT, and the GADIT 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. Some of these questions are:

1. What are the exact types of errors committed within each error category? In other words, how often do skill-based errors involve stick-and-rudder errors, verses attention failures (slips) or memory failures (lapses)?

2. How important is each error type, or how often is each error type the “primary” cause of an accident? For example, 80% of accidents might be associated with skill-based errors, but how often are skill-based errors the “initiating” error or simply the “consequence” of another type of error, such as decision errors?

3. How do the different error types relate to one another, or with other HFACS variables? Are there connections between the categories that, if known, could improve intervention development?

4. Do accidents that occur in different geographical regions or training facilities within the U.S. have different error patterns or trends?

5. What can be done to intervene given the information that is now available, and what more might be done with the additional refined data? Answers to these questions are not available in the database as it currently exists. Therefore, additional fine-grained analyses of the specific human error categories within HFACS are needed to answer these, and other questions that may arise, and to target problem areas within general aviation for future interventions.

<u>Output</u>: The proposed research project, therefore, is in response to these questions and requests made by AFS-800, the ADM JSAT, and the GADIT committees. Specifically, the goal of this project is to perform a comprehensive

and systematic analysis of the individual human causal factors associated with fatal GA accidents. As a joint effort between researchers at the University of Illinois and the FAA's Civil Aerospace Medical Institute, the HFACS framework will be used to perform fine-grained analyses of GA accident data to explore the nature of the underlying human errors associated with these events. The results of these analyses will then be used to map intervention strategies onto different error categories to determine plausible prevention programs for reducing GA accidents. Results will be provided to appropriated FAA officials and committees for consideration. Ultimately, this project will represent the next step in the development of a larger civil aviation safety program whose ultimate goal is to reduce the aviation accident rate through systematic, data-driven investment strategies and objective evaluation of intervention programs.

Regulatory Link:

a. Supports Safer Skies through Areonautical Decision Making (ADM) JSAT

b. AOA (FAA) Strategic Plan (1998-2003) Mission Goal:Safety. Key Strategies "to enable the goal to include identification of root causes of past accidents; and (2) use a more proactive analytical approach, with new data sources, to identify key risk factors and intervene to prevent potential causes of future accidents" (Page 13).

c. FY2001 Performance Plan: Focus Area: Accident Prevention. "Aviation Human Factors to coordinate human factors research, development and based on detailed causal analysis" (Page 2)

d. AVR Performance Plan:Reduce General Aviation fatal accidents (pg 2). Contribute to aviation safety by developing policies,standards, programs, and systems to reduce the number of aviation accidents and incidents related to human factors (pg 9)

Requirement ID: 866

Sponsor Organization: AFS POC: Anne Graham

Requirement Title: Low Visibility and Visual Detection

Funded Requirement:
- FY02: No
- FY03: Yes (late FY)
- FY04: Yes
- FY05: Yes

Requirement Statement: 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 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

Background:
Problem 1: VFR into IMC
Some of the most difficult safety issues currently being addressed by the FAA include accidents in which reduced visibility or failures to visually detect other aircraft or ground-based targets played a major role. In 1989, the “Final Report on an Informal Panel on General Aviation Safety Submitted to J. Lynn Helm” identified VFR into IMC as the leading cause of fatal crashes. Night VFR minimums were increased and other interventions implemented, but the problem still exists. Interventions that focus on improved pilot training concerning weather related decision making are critical in reducing fatalities. A review, of poor

visibility CFIT crashes in Alaska since 1980, indicates that pilots failed to transition to an emergency operation or radio for help. Most appear to be attempting visual flight until impact.

Problem 2: failure to detect targets

Similarly, the midair and ground collision rates involving GA aircraft remains unacceptably high. Reports of collisions both in the air and on the ground indicate that the pilots were typically unaware of decreasing visibility or camouflage effects from the background and report never having seen the target until too late. One prime example is the failure of pilots top detect other airborne traffic. Ironically this situation often happens in clear weather. The major cause of this failure is camouflage or masking effects of high contrast backgrounds, such as snow on mountains, or buildings in an urban landscape (please see examples in accompanying video). One of the primary cues for detection of targets is motion. However motion cues are of little help when targets are on a direct collision course since there is no relative motion in this case. The strategy of frequently changing course direction frequently may improve target detection in these cases.

The current recommended target scanning technique is based on the assumption that target detection always occurs with the central (foveal) area of the retina. This assumption ignores the specialized processing that occurs in the paracentral and peripheral areas of the retina that are optimized to detect transient change (motion and flicker). It is possible that modifications of the recommended scanning techniques to more efficiently utilize motion detection capacities will improve detection when combined with intentional course changes under conditions of target masking. More research on this topic is needed.

Problem3: failure to utilize resources

A lot can be said in favor of the new technologies associated with the Capstone project as well as traffic avoidance systems. However, widespread use of this equipment is most likely to be a long time coming and prohibitively expensive for many GA operators. Additionally, as we saw in the recent Kennedy crash, having sophisticated technologies on board does not assure they will be used properly. The more airman know of the limitations for both man and machine under non-optimal visual conditions the more likely they will avoid the situation or will be prepared to handle it.

Problem 4: pilot education

Despite the seriousness of the current situation, information about physiological and psychological responses to deteriorating weather conditions or reduced visual cues and detection is not widely disseminated. The aviation industry has been primarily focused on how to prevent pilots from entering visibilities below VFR minimums, yet it happens and fatal accidents occur. Basic and applied research with an aim toward improving training practices concerning operations in conditions of reduced visibility and detection is important. More information will help the pilot, who is faced with challenging visual conditions to better cope with this predicament.

Output: Years 1-3
1. There is a lack of data on pilot performance under varying task loads in reduced visibility conditions. Data on this topic could be used to develop advisory circulars or to develop training modules, which would make pilots aware of their limitations and the difficulty of flying and navigating while in reduced visibility at low altitudes and when targets may be efficiently hidden by background conditions. The specific product that is needed is a report that quantifies the relationships between pilot performance, task load (as indicated, for example, by aircraft speed and altitude) and visibility.
2. There is a need to improve pilot decision-making during potential collision and CFIT situations. One common model of pilot decision-making portrays the decision-making process as a continuous loop. On the other hand, in a high task load environment like low altitude and low visibility operations a "discontinuous decision-making" model would most likely be of value. The specific product needed is a report that evaluates currently used poor weather decision models, such as Bensyl, American Journal of Epidemiology, December 2001 or Controlled Flight Into Terrain: A Study of Pilot Perspectives in Alaska, Larry Bailey, Civil Aeromedical
3. Even with advanced display technology, like weather and terrain displays on board, inadequate decision making could result in and an accident. In a low visibility and low altitude environment the man-machine interface is critical. Little is known about the advance technology equipment training and proficiency needed to contend with a VFR into inadvertent IMC situation. The specific product that is needed is a report that specifies inadequate techniques and the techniques that experienced pilots have found to be effective in dealing with these conditions. Information of that nature could be incorporated into pilot training programs, much as current emergency procedures are practiced.
4. It is important to educate pilots as to optimal strategies for avoiding accidents in conditions of reduced visibility and where background terrain or objects interfere with the ability to detect possible targets. The specific product needed is an educational video (or CD ROM) that illustrates the problem of low visibility and target detection and the appropriate strategies for reducing the probability of collision or CFIT. Data from the reports generated by Output 1, 2 and 3 above, would be incorporated into and form the basis of this product.

Regulatory Link: Safer Skies: Goal to Reduce of Fatalities, Reduction of CFIT accidents, Reduction of Weather Related Accidents, and Improving Pilot Decision-Making.

Requirement ID: 869

Sponsor Organization: ACE POC: Frank Bick

Requirement Title: Primary Flight Displays, terrain, overlays/layers, perspective displays

Funded Requirement:
- FY02: No
- FY03: Yes
- FY04: Yes
- FY05:

Requirement Statement: The intent of this research requirement is to identify factors salient to the design and certification of primary flight displays that may contain terrain representations and flight guidance cues and to quantify their effects upon pilot performance (flight technical error, procedural performance, and terrain awareness). Not all of the listed issues will necessarily be addressed by empirical research, particularly where there are extant data pertaining to the question. Issues to be examined include the following: Manner of horizon depiction independently from the terrain to guarantee its availability to the pilot; optimizing format of terrain as a function of phase of flight; providing for deselection of terrain depiction; indications when extreme attitudes place terrain out of view; indications for failed or deselected terrain depiction; effects of variables associated with wire-frame presentations; point of regard (viewing vector); aiding recovery from unknown attitudes; use of pitch ladders; color coding schemes; optimal field of view by task; display aspect ratio; comparison with baseline standard instrumentation; substitution of other display enhancements for HITS-format guidance when terrain depiction present; separation or integration of terrain and flight-path guidance symbology. A summary of extant data will be prepared and empirical research will be used to obtain those data not available in the literature.1443

Background: Recent applications for certification of electronic flight displays have included aircraft attitude instrumentation/primary flight displays that depict perspective terrain as well as basic attitude information. In some cases there are also data for airspeed, altitude, and other flight-performance parameters. The manner in which these data are "integrated" can have a significant effect on pilot performance, particularly if the combining leads to clutter or the obscuration of key data because of inappropriate layering schemes. Data are needed to aid certification personnel in assessing which display formats, if any, will produce acceptable levels of safety in operations using these terrain-inclusive displays. The displays in question are any forms of display (head-down panel-mounted, head-up, head- or helmet-mounted) that are permanently installed in the aircraft and depict terrain or terrain with separate attitude indications as the primary means of assessing aircraft attitude.

The data required include but are not limited to graphical formatting of the terrain for presentation with attitude information (issues involving wire-frame, texture, color, transparency, priority of data), requirements for and formatting of attitude indices separate from the terrain depiction, and workload issues associated with major variations in display format.
There is an ongoing concern about the presentation of command guidance information on primary flight displays, including various forms of flight directors and highway-in-the-sky formats. Applicants for certification of new displays are now looking at using pathway formats for primary guidance, and data are needed by the certification community to determine how the level of safety attainable with these displays compares with that currently attainable with more conventional presentations, and if there are format issues that have critical impacts on pilot performance. Some of these data concerning display format effects are already available, but baseline data for performance with a flight-director display are needed that are directly comparable with those data already collected for pathway-format displays.
An additional concern when using such displays is to what degree the data provided are sufficient for maintaining attitude and altitude awareness. That is, to what extent can the terrain data alone be used as an attitude reference and as a means of maintaining separation from the terrain and obstacles on the terrain? The degree to which the displays provide usable information will directly impact the efficacy of use for recovery from unusual or unknown attitudes and the avoidance of controlled-flight-into-terrain accidents. Although it is expected that the terrain representation will serve as a redundant cue for both attitude and altitude information, reason exists to believe that the pictorial nature of the presentation may make it compelling and that it can and may exert an disproportionate influence over the pilot's interpretation of the overall situation.

Output: The performing activity will determine what factors are the major contributors to significant variations in pilot performance resulting from the use of terrain representations in primary flight displays, assess differences in pilot performance between "baseline" instrumentation and terrain-inclusive presentations for selected representative piloting tasks, and provide a summary of these findings in a form that certification personnel can use to determine the acceptability of displays, based upon human factors/human performance criteria, submitted for certification.
Specific questions to be addressed:
1) How should the horizon be depicted independently from the terrain to guarantee its availability to the pilot?
2) What format of terrain is 'best' as a function of phase of flight? (takeoff, climb, cruise, decent, approach)
3) Should the terrain depiction be selectable, i.e., is a provision for switching off the terrain an enhancement or detriment to safety?
4) If no terrain is visible in the display, what should the indication be?
5) Should the display be wire-frame grid?
6) What is the Point of regard?

7) Should the display have pitch ladder or comparable indication?
8) Color coding of terrain: What scheme should be used if color employed?
9) What is the field of view?
10) What is the aspect ratio?
11) Is a PFD with terrain depiction better than the standard instruments?
12) Do other display enhancements such as a velocity vector or other implementations preclude the need for a highway-in-the-sky depiction for flight-path guidance?
13) Should the synthetic terrain appear behind or be integrated with the PFD attitude and flight-path guidance symbology?

Regulatory Link: The sponsor will use the data to refine guidelines for the certification of PFDs containing terrain depictions and/or perspective graphical flight-path guidance indicators. The data will also be used to generate appropriate guidance documentation (certification check lists, advisory circulars, guidelines for potential applicants, other documents) where applicable.