# Office of the Chief Scientist for Human Factors

# General Aviation Human Factors

## Program Review
## FY04

**Dr. William K. Krebs**

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

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

The following report summarizes projects between October 1st, 2003 and September 30th, 2004. These projects attempt to address requirements identified by the Federal Aviation Administration Flight Standards and Certification offices. The intent of this report is to allow Federal Aviation Administration sponsors to determine whether their requirements have been satisfactorily addressed, allow investigators to receive feedback from Federal Aviation Administration sponsors and other interested parties, and to provide feedback to the ATO-P R&D HF general aviation program manager on the quality of the research program. Basically, this document is a means of holding each group (sponsor, investigator, ATO-P R&D HF program manager) accountable to ensure that the program is successful.

In FY04, the general aviation research program distributed $437,000 contract dollars to performing organizations. In addition, some of these projects received supplemental support from the Civil Aerospace Medical Institute, Oklahoma City, OK.

Address questions or comments to:

William K. Krebs, Ph.D.

# General Aviation Human Factors

## FY04 Funded Projects

# 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

The Human Factors Analysis and Classification System (HFACS) is a theoretically based tool for investigating and analyzing human error associated with accidents and incidents. Previous research performed at both 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 causes of both commercial and general aviation (GA) accidents. These analyses have identified general trends in the types of human factors issues and aircrew errors that have contributed to civil aviation accidents. The next step is to identify the exact nature of these human errors. The purpose of this research effort was to address these questions by performing a fine-grained HFACS analysis of the individual human causal factors associated with GA accidents to assist in the generation of intervention programs.

## INTRODUCTION

Ultimately, most aviation accidents do not happen in isolation, rather, they are the result of a chain of events often culminating with the unsafe acts of aircrew. From Heinrich's (Heinrich, Peterson, & Roos, 1931) axioms of industrial safety, to 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 (Wiegmann & Shappell, 2001c). Reason's (1990) description of active and latent failures within the context of his "Swiss cheese" model of human error has been particularly useful in this regard.

Within his model, Reason describes four levels of human failure, each one influencing the next. According to Reason, *organizational influences* often lead to instances of *unsafe supervision*, which in turn lead to *preconditions for unsafe acts* and ultimately the *unsafe acts of operators*. It is at this latter level, the unsafe acts of operators, that most accident investigations focus.

Unfortunately, while Reason's work forever changed the way aviation and other accident investigators view human error; it was largely theoretical and did not provide the level of detail necessary to apply it in the real world. 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 integrated into the applied setting.

## HFACS

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

### Unsafe Acts of Operators

In general, the unsafe acts of operators (in the case of aviation, the aircrew) can be loosely classified as either errors or violations (Reason, 1990). Errors represent the mental or physical activities of individuals that fail to achieve their intended outcome. Not 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.

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

While errors occur when aircrews are behaving within the rules and regulations implemented by an organization, violations represent the willful disregard for the rules and regulations that govern safe flight. As with errors, there are many ways to distinguish between types of violations. However, two distinct forms are commonly referred to, based upon their etiology. The first, routine violations, tend to be habitual by nature and are often tolerated by the governing authority. The second type, exceptional violations, appear as isolated departures from authority not necessarily characteristic of an individual's behavior nor condoned by management.

## PURPOSE

The HFACS framework was originally developed for the U.S. Navy and Marine Corps as an accident investigation and data analysis tool (Shappell & Wiegmann, 2000; 2001; Wiegmann & Shappell, 2003). Since it's development however, other organizations such as the FAA 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 at both the University of Illinois and the Civil Aerospace Medical Institute (CAMI) have been highly successful and have shown that HFACS can be reliably and effectively used to analyze the underlying human causes of both commercial and general aviation accidents (Wiegmann & Shappell, 2003). 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 (Shappell & Wiegmann, 2003; Wiegmann & Shappell, 2001a; 2001b).

The FAA's General Aviation & Commercial Division (AFS-800) within the Flight Standards Service and the Small Airplane Directorate (ACE-100) have acknowledged the added value and insights gleaned from these HFACS analyses. Likewise, HFACS was cited by the Aeronautical Decision Making (ADM) Joint Safety Analysis Team (JSAT) and the General Aviation Data Improvement Team (GADIT) as particularly useful in identifying the human error component of aviation accidents.

To date, however, the analyses using HFACS have generally been performed at a global level, leaving several questions unanswered concerning the underlying nature and prevalence of different error types. As a result, AFS-800, ACE-100, the ADM JSAT, and the GADIT have directly requested that additional analyses be conducted to answer specific questions about the exact nature of the human errors identified, particularly within the context of GA.

### Previous Findings

For a complete accounting of this work, please see the FY02 and FY03 Annual Reports. In sum however, previous research performed at the University of Illinois and CAMI over the past two years has revealed that roughly 80% of GA accidents are associated with skill-based errors, followed by decision errors (roughly 30%), violations (16%), and perceptual errors (5%; Figure 1). Equally important, the trends for the unsafe acts across the years have not changed.

Moreover, upon examination of the fatal and non-fatal aircrew error data during the years of this study, the only difference between the human error categories was for violations. That is, fatal accidents were four times more likely to be associated with a violation than non-fatal accidents.

The pattern of results was similar when the data were examined for the "initiating" or seminal event in the accident chain.[1] Indeed, nearly 61% (n = 8,838) of all accidents began with a skill-based error. In contrast, roughly 19% (n = 2,729) of the accidents examined began with a decision error, 8% (n = 1,180) began with a violation and only 4% (n = 564) began with a perceptual error. The remaining 8% (n = 1,125) were associated with a seminal event other than an unsafe act (e.g., a precondition for an unsafe act, such as an adverse physiological state).

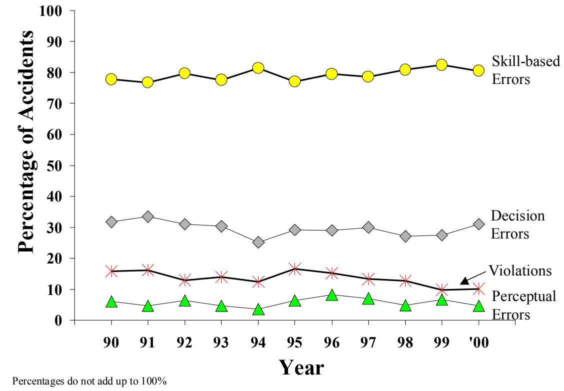



Figure 1. Percentage of accidents by error category by year.

When comparing fatal versus non-fatal seminal errors, what differences did occur (i.e., skill-based and violations) remained relatively constant across the years of this study. Furthermore, the differences were in opposite directions with a higher percentage of fatal than non-fatal accidents associated with violations and a higher percentage of non-fatal than fatal accidents associated with skill-based errors.

### *FY04 Research Effort*

The current research effort focused on the following questions that had also been posed by AFS-800, ACE-100, the ADM JSAT, and GADIT.

***Question 1:*** *What are the exact types of errors committed within each error category?*

***Question 2:*** *Do the types of errors committed within each error category differ across accident severity?*

***Question 3:*** *Do the types of errors committed within each error category differ between seminal vs. non-seminal unsafe acts?*

## METHOD

### Data

General aviation accident data from calendar years 1990-2000 was obtained from databases maintained by the National Transportation Safety Board (NTSB) and the FAA's National Aviation Safety Data Analysis Center (NASDAC). For analysis purposes, we selected only those accident reports that were classified "final" at the time this report was written, since

[1] Note that unlike the previous analysis where the percentages will add up to more than 100% because there is typically more than one cause factor per accident, these percentages will add up to 100%, since there can only be one "seminal" human causal factor.

only those reports contain the causal factors associated with the accident.

We further eliminated those accidents that were classified as having "undetermined causes," and those attributed to sabotage, suicide, or criminal activity (e.g., stolen aircraft). When the data were parsed in this manner, we were left with only those GA "accidents" for which causal factors had been "determined" and released by the NTSB.

The data were then culled further to include only those accidents that involved powered GA aircraft (i.e., airplanes, helicopters, and gyrocopters). Finally, since we were interested only in aircrew error, we excluded accidents in which no aircrew-related unsafe act was considered causal or contributory to the accident. In the end, 14,436 accidents involving over 25,000 aircrew causal factors were included and submitted to further analyses using the HFACS framework.

*Causal Factor Classification using HFACS*

Seven GA pilots were recruited from the Oklahoma City area as subject matter experts (SMEs). All were certified flight instructors with a minimum of 1,000 flight hours in GA aircraft at the time they were recruited. Each pilot was provided roughly 16 hours of training on the HFACS framework. After training, the SMEs 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 SMEs were instructed to classify each human causal factor identified by the NTSB using the HFACS framework. After the 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 SMEs were instructed 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.

*Human Factors Quality Assurance*

General aviation pilots are not SMEs in the domains of psychology or human factors, and therefore, they may not fully understand the theoretical underpinnings associated with the various error types within the HFACS framework. Hence, pilots might classify human error data somewhat differently than SMEs in human factors. Still, pilots in this study were trained on HFACS, which did give them some level of expertise when assessing human error.

Nonetheless, to be sure that the SMEs had grasped the psychological aspects underlying human error and HFACS, three additional SMEs with expertise in human factors/aviation psychology examined each HFACS classification that the pilot SMEs had assigned to a given human cause factor. Essentially, the human factors SMEs were ensuring that the pilots understood the error analysis process and did not code causal factors like spatial disorientation as a decision error, or exhibit any other misunderstandings of the HFACS model. To aid in the process, descriptive statistics were used to identify outliers in the data, after which the corresponding NTSB report was obtained. The reports were then independently reviewed by a minimum of two human factors (HF) SMEs for agreement with the previous codes. After the HF SMEs came to a consensus, the codes were either changed in the database or left as the pilot SMEs originally coded them. In the end, less than 4% of all causal factors were modified during the human factors quality assurance process.

## RESULTS

Just knowing that skill-based errors (or any other type of error) are a major concern does not provide safety professionals sufficient detail to do anything about it. What was needed was a fine-grained analysis of the specific types of errors within each HFACS causal category, so that targeted interventions can be developed. With this in mind, we compared each HFACS classification with the NTSB's causal factor designation.

To aid in the presentation of the data, we will examine the fine-grained analysis for each type of unsafe act separately. Included in the results will be the ***"top 5"*** human causal factors overall, across accident severity, and seminal events.

***Skill-based errors.*** The most frequently occurring human error categories within skill-based errors are presented in Table 1. As can be seen, nearly 12% of all skill-based errors involved errors in maintaining direction control, followed by airspeed (10.63%), stall/spin (7.77%), aircraft control (7.62%) and errors associated with compensating for wind conditions (6.18%). Together, these five cause factors accounted for nearly one half of all the skill-based errors in the database. Additionally, the types and frequencies of skill-based errors coded as fatal/non fatal and seminal events are also shown in Table 1. The percentage of skill-based errors involving stall/spin, airspeed, and aircraft control were greater for fatal than non-fatal accidents. In contrast, causal factors such as directional control and compensation for wind conditions were rarely associated with fatal accidents.

Such findings make sense when one considers that errors leading to a stall/spin, as well as airspeed and control of the aircraft in the air typically happen at altitude, making survival less likely. In contrast, errors controlling the aircraft on the ground (such as ground loops) and compensation for winds (typically seen during cross-wind landings), while dangerous, don't necessarily result in fatalities.

*Decision Errors.* Table 2 presents the most frequently occurring decision errors. Improper in-flight planning tops the list, contributing to roughly 18% of all decision errors. The remaining decision errors, such as preflight planning/decision errors (8.94%), fuel management (8.73%), poor selection of terrain for takeoff/landing/taxi (7.85%), and go-around decisions (6.03), all occurred at approximately the same frequencies. Combined, these five causal categories accounted for roughly half (49.89%) of all decision errors in the database. It should be noted, individual factors related to weather-related decision making did not reach the top of the list (e.g., weather evaluation, flight into adverse weather, and inadvertent VFR flight into IMC). However, when combined, they did constitute a significant portion of the factors related to decision- making (6%).

Table 2 also presents the types and frequencies of decision errors for fatal/non fatal and seminal events. As indicated, the categories in-flight planning and planning/decision making on the ground tended to be associated more often with fatal than non-fatal accidents. Whereas the categories unsuitable terrain, go around, and fuel management were associated more often with non-fatal accidents. This pattern was generally consistent for the overall data, as well as within seminal events.

*Perceptual errors.* A review of accident causes and factors coded as perceptual errors revealed that misjudging distance was most common, accounting for over a quarter of all perceptual errors (26.4%; see Table 3). The next highest was flare (22.5%), followed by misperceiving altitude (11.4%), misjudging clearance (7.0%) and visual/aural perception (5.1%). Together these errors accounted for nearly three quarters of all perceptual errors in the database.

The types and frequencies of perceptual errors as they occurred within fatal/non-fatal accidents are also shown in Table 3. There was very little difference in the percentage of fatal and non-fatal accidents associated with any particular type of perceptual error. The only exception appears to be perceptual errors related to performing the flare, which in most cases is associated more with non-fatal than fatal accidents.

*Violations.* The *top five* violations are presented in Table 4. Analysis of the fundamental types of unsafe acts that are included within the violations categories reveals that the most common violation involved visual flight rules (VFR) flight into instrument meteorological conditions (IMC) (15.5%) and not following known procedures or directives (10.9%). The remaining top violations included operating aircraft with known deficiencies (9.9%), performing hazardous maneuvers, such as low altitude flight or buzzing (8.7%), and flight into adverse weather (8.5%). Together, these five variables accounted for over half of all violations in the database.

The types and frequencies of violations for fatal/non-fatal and seminal events are also presented in Table 4. As indicated, the categories VFR flight into IMC, hazardous maneuver, and flight into known adverse weather were much more likely to be fatal than non-fatal, both overall and for seminal events only. This pattern is consistent with the observation that accidents involving violations of the rules are, in general, more likely to be fatal.

Table 1. *Five Most Frequent Skill-based Error Categories for Fatal and Non-fatal Accidents.*

| ERROR CATEGORY | OVERALL Frequency (%) | | | SEMINAL Frequency (%) | | |
|---|---|---|---|---|---|---|
| | Fatal | Non-fatal | Total | Fatal | Non-fatal | Total |
| **Directional Control** | 20 (0.50) | 2018 (15.2) | 2038 (11.8) | 9 (0.57) | 1326 (17.5) | 1335 (14.6) |
| **Airspeed** | 713 (17.9) | 1127 (8.5) | 1840 (10.6) | 302 (19.2) | 605 (8.0) | 907 (9.9) |
| **Stall/Spin** | 592 (14.9) | 753 (5.7) | 1345 (7.8) | 84 (5.3) | 144 (1.9) | 228 (2.5) |
| **Aircraft Control** | 654 (16.5) | 665 (5.0) | 1319 (7.6) | 311 (19.8) | 429 (5.7) | 740 (8.1) |
| **Compensation for winds** | 23 (0.6) | 1046 (6.2) | 1069 (6.2) | 12 (0.8 | 859 (11.4) | 871 (9.5) |

Table 2. *Five Most Frequent Decision Error Categories for Fatal and Non-fatal Accidents.*

| ERROR CATEGORY | OVERALL Frequency (%) | | | SEMINAL Frequency (%) | | |
|---|---|---|---|---|---|---|
| | Fatal | Non-fatal | Total | Fatal | Non-fatal | Total |
| **In-flight Planning** | 268 (22.9) | 683 (17.0) | 951 (18.3) | 133 (22.6) | 427 (19.8) | 560 (20.4) |
| **Planning/Decision-making on the Ground** | 115 (9.8) | 349 (8.7) | 464 (8.9) | 89 (15.1) | 284 (13.1) | 373 (13.6) |
| **Fuel Management** | 40 (3.4) | 413 (10.3) | 453 (8.7) | 20 (3.4) | 252 (11.7) | 272 (9.9) |
| **Unsuitable Terrain Selection** | 16 (1.4) | 391 (9.8) | 407 (7.8) | 5 (.85) | 284 (13.1) | 289 (10.5) |
| **Go Around** | 22 (1.9) | 291 (7.3) | 313 (6.0) | 5 (.85) | 70 (3.2) | 75 (2.7) |

Table 3. *Five Most Frequent Perceptual Error Categories for Fatal and Non-fatal Accidents.*

| ERROR CATEGORY | OVERALL Frequency (%) | | | SEMINAL Frequency (%) | | |
|---|---|---|---|---|---|---|
| | Fatal | Non-fatal | Total | Fatal | Non-fatal | Total |
| **Distance** | 26 (17.8) | 233 (27.7) | 259 (26.4) | 23 (33.8) | 135 (26.5) | 158 (27.4) |
| **Flare** | 5 (3.4) | 217 (25.8) | 222 (22.5) | 4 (5.9) | 163 (32.0) | 167 (28.9) |
| **Altitude** | 22 (15.1) | 91 (10.8) | 113 (11.4) | 9 (13.2) | 51 (10.0) | 60 (10.4) |
| **Clearance** | 18 (12.3) | 51 (6.1) | 69 (7.0) | 14 (20.6) | 41 (8.1) | 55 (9.5) |
| **Visual/Aural Perception** | 15 (9.6) | 36 (4.2) | 50 (5.1) | 3 (4.4) | 5 (1.0) | 8 (1.4) |

Table 4. *Five Most Frequent Violations for Fatal and Non-fatal Accidents.*

| ERROR CATEGORY | OVERALL Frequency (%) | | | SEMINAL Frequency (%) | | |
|---|---|---|---|---|---|---|
| | Fatal | Non-fatal | Total | Fatal | Non-fatal | Total |
| **VFR Flight into IMC** | 305 (25.8) | 53 (4.7) | 358 (15.5) | 182 (30.5) | 29 (5.2) | 211 (25.8) |
| **Procedures/Directives Not Followed** | 75 (6.3) | 176 (15.6) | 251 (10.9) | 37 (6.2) | 109 (19.6) | 146 (12.7) |
| **Operating Aircraft with Known Deficiencies** | 61 (5.2) | 168 (14.9) | 229 (9.9) | 27 (4.5) | 97 (17.4) | 124 (10.8) |
| **Hazardous Maneuver** | 154 (13.0) | 47 (4.2) | 201 (8.7) | 83 (13.9) | 24 (13.9) | 107 (9.3) |
| **Flight into Known Adverse Weather** | 135 (11.4) | 61 (5.4) | 196 (8.5) | 85 (14.3) | 41 (7.4) | 126 (10.9) |

## DISCUSSION

The high level of safety currently achieved within aviation should not obscure the fact that many aviation accidents are preventable. It is important to realize that safety measures and defenses currently in place in GA may be inadequate, circumvented, or perhaps ignored, and that the intervention strategies aimed at reducing the occurrence or consequences of human error may not be as effective as possible.

The present study of GA accidents examined literally thousands of unsafe acts committed by pilots, perhaps suggesting that, correspondingly, there are literally thousands of unique ways to crash an airplane. The results of this study, however, demonstrate that accidents that may appear to be unique on their surface can be reliably grouped based upon underlying cognitive mechanisms of pilot errors. By applying HFACS, a theoretically based model of human error, we were able to highlight several human error trends and identify the categories of unsafe acts that contribute to both fatal and non-fatal GA accidents. Ideally, data such as this will result in more data-driven intervention efforts being developed and implemented.

## REFERENCES

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

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

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. (2001a). Human error analysis of commercial aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Proceedings of the Eleventh International Symposium on Aviation Psychology*, Ohio State University.

Wiegmann, D. and Shappell, S. (2001b). 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. (2001c). Human error perspectives in aviation. *International Journal of Aviation Psychology, 11*(4), 341-357.

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.

## 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 preliminary trend indicates 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 was 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

In the initial proposal a total of 180 pilots (30 in each of the 6 groups) were scheduled to participate in the study. Due to funding reductions in the second and third years, the number of pilots in the study was reduced to a total of 120 pilots (20 subjects in each group). Due to the elimination of FY 2005 funding the best case of the number of subjects currently ranges between 16 and 20. The subjects are University of Illinois, Institute of Aviation private pilot students, who are enrolled in the Institute's instrument program. To date 91 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 for the four groups ( Airplane, PCATD, FTD 5 and 10 groups) that received only prior training on instrument tasks . 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, all students, a total of 124, have completed and taken the final check ride the AVI 130 Basic Instruments course. This is an increase from 65 from last year's report. Table 1 shows the results of the check ride for the six groups for the fall 2002, spring, summer and fall 2003, and spring 2004 semesters. A total of 75 students passed the check ride on the first attempt and 49 students passed on the second attempt. Nine students have been recommended for a remedial course, AVI 102. The total dual flight time to completion for AVI 130 (the basic instrument course) for the six groups is shown in Table 1 and in Figure 1. The average dual flight time to course completion for the Airplane Group was greater than the average time for each of the five experimental groups who had prior training in the PCATD or the FTD. The Airplane group required 22.35 hours of dual to complete the course while the five experimental groups, after prior training in the PCATD or the FTD, the dual flight time in the airplane ranged between 18.31and 20.87 hours. A total of 95 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 48 students passed the check ride on the first attempt and 40 students passed on the second attempt. There were no students recommended for remedial training (AVI 102) in the summer 2004 session. The total dual flight time to completion for the six groups for the advance instrument course (AVI 140) is shown in Table 2 and in Figure 2. The average course completion time for the Airplane Group is greater for each of the five experimental groups who had prior training in the PCATD or the FTD. The Airplane group required 26.02 hours of dual to complete the course while the hours to completion for the five experimental groups ranged from the dual flight time in the airplane ranged between 25.77 - 20.11 hours after prior training in the PCATD or the FTD. Statistical analyses based on current data indicate no significant differences between the three experimental groups that received prior training on instrument tasks and the Control group.

## DISCUSSION

The trend from 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 but the limited number of subjects has prevented this trend from reaching statistical significance. With the limited number of subjects and the current variability among subjects the power is low. If this trend is confirmed this study will systematically replicate the findings of Taylor et al. (1996, 1999) that PCATDs are useful to teach instrument tasks to private pilots. As a result of prior training in an FTD and a

PCATD time to the stage check in AVI 130 and to the instrument rating flight check was less for 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. The current project is evaluating the effectiveness of using FTDs for 5 and 10 hours of cross-country flight. The data thus far indicate that additional FTD time can be effectively used during 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, Flight Standard Service, General Aviation & Commercial Division. Dennis Beringer serves as the COTR. Views expressed herein do not necessarily represent official FAA positions. We express our appreciation to Ms. Sybil Phillips for invaluable assistance with flight operations and with student management. Mr. Bill Jones, Mr. David Boyd, and Don Talleur served as check pilots. We also thank the Institute of Aviation 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, Spring, Summer, Fall 2003 and Spring 2004)**

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

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

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

| | Airplane Only | PCATD 5.00 | Frasca 5.00 | Frasca 10.00 | Frasca 15.00 | Frasca 20.00 |
|---|---|---|---|---|---|---|
| Number of Students | 17 | 17 | 17 | 15 | 13 | 16 |
| % First Flight Pass Rate | 47.06 (N=8) | 52.94 (N=9) | 52.94 (N=9) | 40.00 (N=6) | 46.15 (N=6) | 62.50 (N=10) |
| % Second Flight Pass Rate | 100.00 (N=9) | 75.00 (N=6) | 100.00 (N=6) | 88.89 (N=8) | 100.00 (N=7) | 66.67 (N=4) |
| Students Recommended 102 | 2 | 3 | 4 | 2 | 4 | 2 |
| Total Dual to Completion | 26.02 (N=17) | 25.77 (N=16) | 24.55 (N=15) | 23.78 (N=15) | 22.18 (N=13) | 20.11 (N=15) |
| Variance Total Dual to Completion | 15.10 | 6.43 | 7.74 | 8.87 | 11.25 | 11.30 |

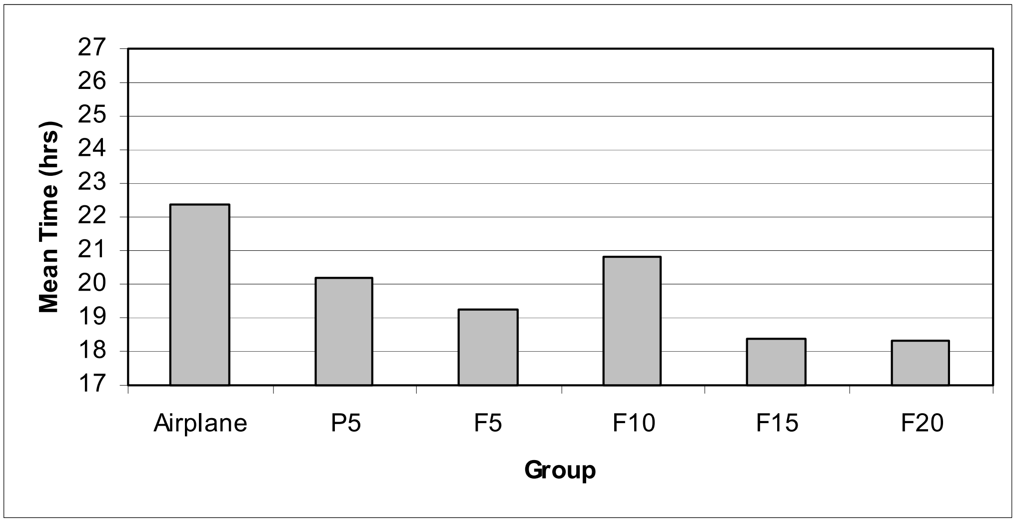



**Figure 1. Flight Lesson 45 Statistics (Fall, 2002, Spring, Summer, Fall 2003, and Spring 2004)**

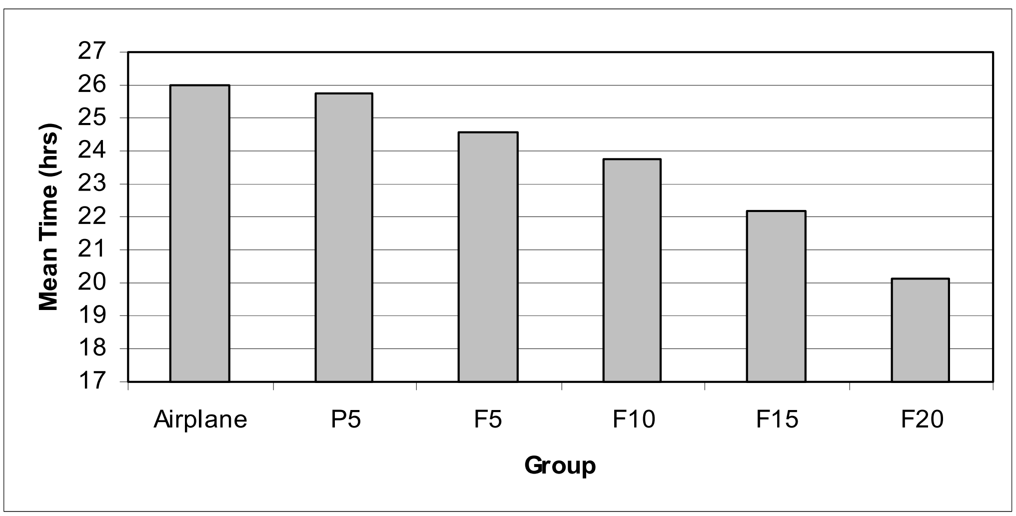



**Figure 2. Flight Lesson 60 Statistics (Spring, Summer, Fall, 2003, Spring, Summer 2004)**

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

This project evaluated the effectiveness of a personal computer aviation training device (PCATD), a flight training device (FTD) and an airplane in conducting an instrument proficiency check (IPC). The study compared the performance of pilots receiving an IPC in a PCATD, in a FTD and in an airplane (IPC #1) with performance on an IPC in an airplane (IPC #2). Chi-square tests were used to analyze the IPC #1 and IPC #2 data to determine whether the treatment (assignment to group) had an effect on the pass/fail ratio for the IPC #1 and IPC #2 flights respectively. The treatment effect on the IPC #1 pass/fail ratios was not statistically significant. Neither was the treatment effect statistically significant on the IPC #2 pass/fail ratio. A series of planned-comparison tests were performed between and among the experimental groups. The first comparison evaluated the performance of the PCATD group on IPC #2 with the Airplane. The next comparison evaluated the performance of the PCATD group on IPC #2 with the FTD Group. Neither of the comparisons was significant. The final comparison, which was not significant, evaluated the performance of the Airplane group on IPC #2 with the Frasca group.
.

## 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, 2003). 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 was 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 investigated 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 permitted 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). Most of the participating pilots were instrument current but a few fall into one of three other categories of instrument currency: (1) within one year of currency, (2) outside of one year of currency but within two years of currency, and (3) outside two years but within five years of currency. All participants received 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. 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.

**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 were 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 received a baseline IPC flight in 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 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, and no additional subjects of this currency status were added to the project. Table 1 depicts the experimental design.

Table 1.
*Experimental Design*

| GROUP | Fam. Flight | Initial IPC flight (IPC#1) | Final IPC flight (IPC#2) |
|---|---|---|---|
| Airplane | In Airplane<br>In Frasca<br>In Elite | IPC flight in Sundowner | IPC flight in Sundowner |
| Frasca | In Airplane<br>In Frasca<br>In Elite | IPC flight in Frasca | IPC flight in Sundowner |
| PCATD | In Airplane<br>In Frasca<br>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 used 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.

## RESULTS

All 75 subjects have completed IPC #1 and IPC #2 The pass/ fail rates by group for IPC #1 and IPC #2 are shown in Table 2.

Table 2.
*Pass/Fail rates by group*

| IPC#1 | | | | | |
|---|---|---|---|---|---|
| Group | N | Pass | (%) | Fail | (%) |
| Aircraft | 25 | 6 | (24) | 19 | (76) |
| FTD | 25 | 9 | (36) | 16 | (64) |
| PCATD | 25 | 9 | (36) | 16 | (62) |
| Total | 75 | 24 | (32) | 51 | (68) |

| IPC#2 | | | | | |
|---|---|---|---|---|---|
| Group | N | Pass | (%) | Fail | (%) |
| Aircraft | 25 | 13 | (52) | 12 | (48) |
| FTD | 25 | 14 | (56) | 11 | (44) |
| PCATD | 25 | 15 | (60) | 10 | (40) |
| Total | 75 | 42 | (56) | 33 | (44) |

Table 2 presents the number and percentage of pilots that passed/failed IPC #1 and IPC #2 for each of the three experimental groups and for the total subjects. Figures 1 and 2 shows the differences between pass rates for the three groups for IPC #1 and IPC #2, respectively. Inspection of Figures 1 and 2 indicate few differences between groups for the number of participants who passed IPC #1 and IPC #2. A total of 24 of 75 subjects (32%) passed the IPC #1 flight in the airplane, FTD and PCATD and a total of 42 of 75 subjects (56%) passed the IPC #2 flight. Chi-square tests were used to analyze the IPC #1 and IPC#2 data to determine whether the treatment (assignment to group) had an effect on the pass/fail ratio for the IPC#1 and IPC#2 flights respectively. The treatment effect on the IPC #1 pass/fail ratios was not statistically significant, $\chi^2$ (2, N=75) = 0.32, p = 0.85. Neither was the treatment effect statistically significant on the IPC #2 pass/fail ratio, $\chi^2$ (2, N=75) = 1.1, p = 0.58.

A series of planned-comparison tests were performed between and among the experimental groups. The first comparison evaluated the performance of the PCATD group on IPC #2 with the Aircraft group, $\chi^2$ (2, N=50) = .32, p > 0.10. The next comparison evaluated the performance of the PCATD group on IPC #2 with the FTD Group, $\chi^2$ (2, N=50) = 0.08, p > 0.10. Neither of the comparisons was significant. The final comparison, which was not significant, evaluated the performance of the Aircraft group on IPC #2 with the Frasca group, $\chi^2$ (2, N=50) = 0.08, p > 0.10.

The pass/fail rates by currency status are shown in Table 3. A total of 53 current pilots took IPC #1 and 19 passed (36%) while 34 failed (64%). Of the 53 current pilots taking IPC #2 and 30 passed (57%) while 23 failed (43%).

Table 3.
*Pass/Fail rates by currency*

| IPC #1 | | | | | |
|---|---|---|---|---|---|
| Currency | N | Pass | (%) | Fail | (%) |
| Current | 53 | 19 | (36) | 34 | (64) |
| Within 1 year | 7 | 2 | (29) | 5 | (71) |
| Within 1-2 years | 1 | 1 | (100) | 0 | (0) |
| 2-5 years | 14 | 2 | (14) | 12 | (86) |

| IPC #2 | | | | | |
|---|---|---|---|---|---|
| Currency | N | Pass | %) | Fail | (%) |
| Current | 53 | 30 | (57) | 23 | (43) |
| Within 1 year | 7 | 6 | (86) | 1 | (14) |
| Within 1-2 years | 1 | 1 | (100) | 0 | (0) |
| 2-5 years | 14 | 5 | (36) | 9 | (64) |

Analysis of the change of performance that took place between the IPC #1 and IPC #2 flights was made in order to understand the effectiveness of the three devices in conducting IPCs. It was expected that performance on IPC #1 would be a good predictor of performance on IPC#2. Table 4 shows a comparison of the pass/fail rates for IPC #1 and IPC #2. Of the 24 participants who passed IPC #1 only 14 also passed IPC #2 (58%), and of the 51 participants who failed IPC #1 only 23 (45%) subsequently failed IPC #2 (a total of 37). Twenty-eight participants, who failed IPC #1 subsequently passed IPC #2 and 10 of the participants who passed IPC #1 subsequently, failed IPC #2 (a total of 38). Therefore, performance on IPC #1 predicted the performance on IPC# 2 only at the chance level. Indeed, the McNemar change in performance analysis between IPC #1 and IPC #2 for all participants was significant; $\chi^2$ (1, $N$ = 75) = 8.53, $p$ < .005.

Table 4.
*IPC #1 vs. IPC #2 Pass/Fail*

| | | IPC#2 | | |
|---|---|---|---|---|
| | | Pass | Fail | Total |
| IPC#1 | Pass | 14 | 10 | 24 |
| | Fail | 28 | 23 | 51 |
| | Total | 42 | 33 | 75 |

## DISCUSSION

This study has demonstrated that there are no significant differences in performance by instrument pilots on an IPC given in either a PCATD, and FTD or an airplane. No significant difference was found on IPC #1 among the three groups, which indicates that the participants performed the same regardless of the device in which they had the IPC. In addition there was no significant difference on IPC #2 indicating that the device in which the participants had IPC #1 had no influence on their performance on IPC #2 in the airplane. The planned comparisons showed that performance on IPC #2 of the PCATD group was statistically indistinguishable from both the airplane and the FTD groups. In addition, there was no difference in performance between the aircraft and the FTD groups. These findings present compelling evidence that the FAA should permit the use of PCATDs to give IPCs.

It was expected that performance on IPC #1 would be a good predictor of performance on IPC#2. A comparison of the pass/fail rates for IPC #1 and IPC #2 indicated that the performance on the baseline IPC was not a good predictor of performance on the final IPC. Only 58 percent of the participants who passed IPC #1 also passed IPC #2 and only 45 percent of the participants who failed IPC #1 also failed IPC #2. Only 49 percent of the participants either passed both tests or failed both tests, while 51 percent of the participants passed IPC #1 and failed IPC #2 or failed IPC #1 and passed IPC #2. Therefore performance on IPC #1 predicts performance on a second IPC at a chance level.

The McNemar change in performance between IPC #1 and IPC #2 for all participants was significant but the comparisons for the individual three groups were not significant. Some of the failures may be related to a lack of familiarity with the PCATD, the FTD and the Sundowner airplane, since few of the participants had flown either of the devices prior to the study. The familiarization flights in each of the devices were expected to provide sufficient familiarity with the devices to eliminate the problem but apparently failed to do so. It is possible that additional familiarity with instrument flying in each device, in addition to the VFR familiarization, was needed. The former was not done in order to minimize a possible training effect on group assignment.

Of the 53 participants who were instrument current, only 19 (36 %) passed IPC #1. The earlier study by Taylor et al. (2001) and Talleur et al. (2003) showed that 42 % of the instrument current pilots passed the initial IPC. The results from the current study are only slightly worse in this regard than those from earlier studies. In addition, most of the participants tested in the previous study had not taken an IPC after the test was standardized to include required maneuvers (thereby increasing the difficulty of the IPC test). This finding raises questions concerning the relationship between instrument currency and instrument proficiency. Less than half of the participants were able to demonstrate instrument proficiency in an IPC in the airplane. This suggests the need for the FAA to consider changing the recency of experience requirements for instrument currency. Taylor et al. (2001) made the same observation and the current study reinforces the concern that currency rules are inadequate for instrument pilots to maintain proficiency. As Taylor et al. (2001) suggested, an alternative approach would be to require a periodic IPC to demonstrate instrument proficiency in addition to the current currency requirements.

## ACKNOWLEDGEMENTS

This work is supported under Federal Aviation Administration (FAA) Cooperative Agreement 2001-G-037. The study was sponsored by FAA Headquarters Flight Standards Service, General Aviation and Commercial Division. 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

Talleur, D. A., Taylor, H. L., Emanuel, T. W., Jr., Rantanen, E, M., & Bradshaw, G. L. (2003). Personal Computer Aviation Training Devices: Their Effectiveness for Maintaining Instrument Currency. *The International Journal of Aviation Psychology, 13(4), 387-399.*

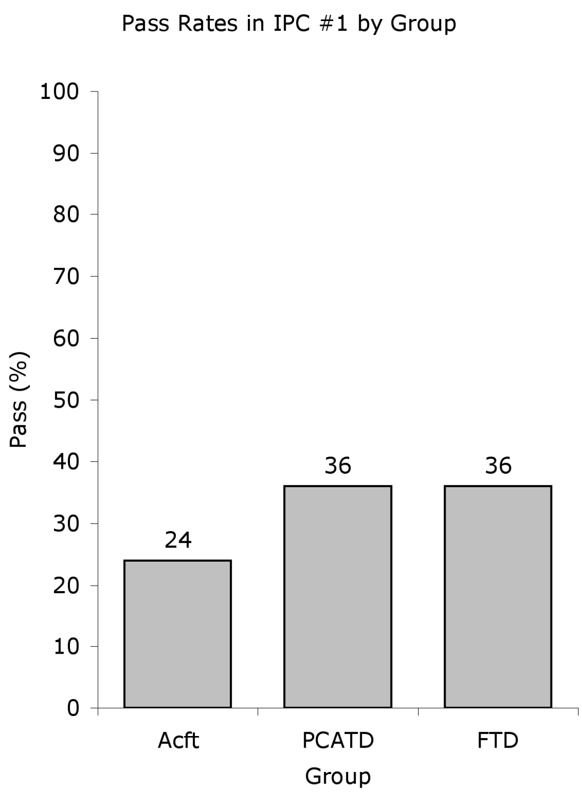
Taylor, H. L., Talleur, D. A., Bradshaw, G. L., Emanuel, T. W., Rantanen, E. M., Hulin, C. L. & Lendrum, L. (2001). *Effectiveness of Personal Computers to meet recency of experience requirements*. (Technical Report ARL-01-6/FAA-01-1). Savoy, IL. University of Illinois, Aviation Research Laboratory

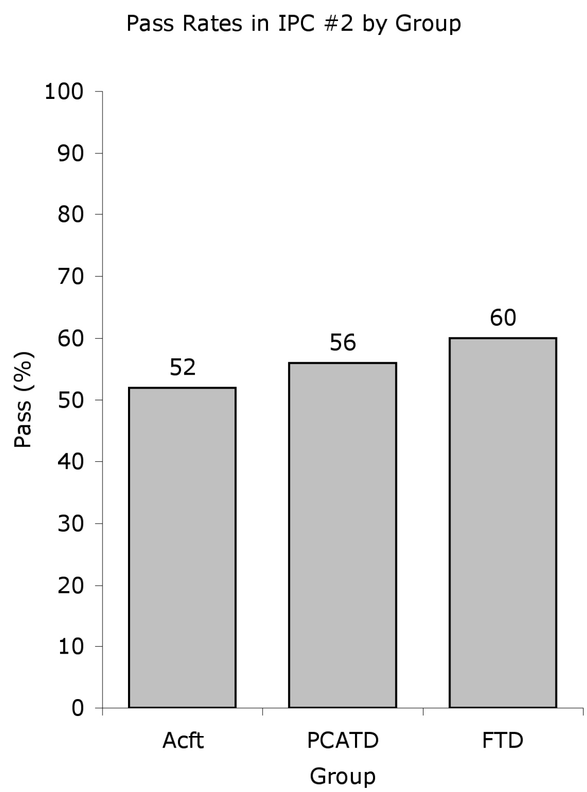
Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D. A., Emanuel, T. W. & Phillips, S. I. (1996). *Transfer of training effectiveness of personal computer-based aviation training devices* (Technical Report ARL-96-3/FAA-96-2). Savoy, IL. University of Illinois, Aviation Research Laboratory.

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.




*Figures 1 and 2.* Pass rates in IPC #1 and IPC #2 by experimental group.

# A Summary of Unmanned Aircraft Accident/Incident Data: Human Factors Implications

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

## Abstract

A review and analysis of unmanned aircraft (UA) accident data was conducted to identify important human factors issues related to their use. UA accident data were collected from the U.S. Army, Navy, and Air Force. The percentage of involvement of human factors issues varied across aircraft from 21% to 68%. For most of the aircraft systems, electromechanical failure was more of a causal factor than human error. One critical finding from an analysis of the data is that each of the fielded systems is very different, leading to different kinds of accidents and different human factors issues. A second finding is that many of the accidents that have occurred could have been anticipated through an analysis of the user interfaces employed and procedures implemented for their use. The current paper summarizes the various human factors issues related to the accidents.

## Introduction

The review and analysis of unmanned aircraft (UA) accident data can assist researchers in identifying important human factors issues related to their use. The most reliable source for UA accident data currently is the military. The military has a relatively long history of UA use and has always been diligent in accurately recording information pertaining to accidents/incidents. The purpose of this research was to review all currently available information on UA accidents and identify human error aspects in those accidents and what human factors issues are most involved.

Two primary sources of accident information were collected from the U.S. Army. The first was a summary of 56 UA accidents produced by the U.S. Army Aeromedical Research Laboratory and obtained from the U.S. Army Risk Management Information System (RMIS). The second was a direct query of the RMIS system of all UA accidents that occurred between January 1986 and June 2004. A total of 74 accidents were identified, the earliest of which occurred on March 2, 1989, and the latest on April 30, 2004.

Information regarding UA accidents for the U.S. Navy was collected from the Naval Safety Center. A summary of 239 UA mishaps occurring between 1986 and 2002 was received from the Naval Safety Center in Pensacola, FL (Kordeen Kor, personal communication).

Air Force accident/mishap information was collected from the Air Force Judge Advocate General's Corps Web site, http://usaf.aib.law.af.mil/. A total of 15 Class-A UA mishaps were retrieved from the Web site, covering the dates from December 6, 1999, to December 11, 2003. In addition, a complete accident investigation board report was received.

Classification of the accident data was a two-step process. In the first step, accidents were classified into the categories of human factors, maintenance, aircraft, and unknown. Accidents could be classified into more than one category. In the second step, those accidents classified as human factors-related were classified according to specific human factors issues of alerts/alarms, display design, procedural error, skill-based error, or other. Classification was based on the stated causal factors in the reports, the opinion of safety center personnel, and personal judgment of the author.

## Results

There are 5 primary military UA in service currently. The U.S. Army's Hunter and Shadow, the U.S. Navy's Pioneer, and the U. S. Air Force's Predator and Global Hawk. Other systems are being developed and have undergone testing, such as the Mariner system for the U.S. Coast Guard and U.S. Navy but sufficient accident data do not exist to warrant separate analyses of these airframes.

### Hunter

The Hunter takes off and lands using an external pilot (EP), standing next to the runway in visual contact with the aircraft, and operating a controller that is very similar to ones used by radio-controlled aircraft hobbyists. After takeoff and climb out, control of the aircraft is transferred to an internal pilot (IP), operating from a ground control station (GCS). The IP controls the Hunter in a more automated fashion, by selecting an altitude, heading, and airspeed for the aircraft using a set of knobs located within the GCS. For landing, control of the aircraft is transferred from the GCS back to an EP. A hook located below the aircraft is used to snag the aircraft on a set of arresting cables positioned across the runway.

Data from the Hunter program indicated that 15 of the 32 accidents (47%) had one or more human factors issues associated with them. Figure 1 shows the major causal categories for Hunter accidents. Note that the percentages add to more than 100% because some of the accidents were classified into more than one category.

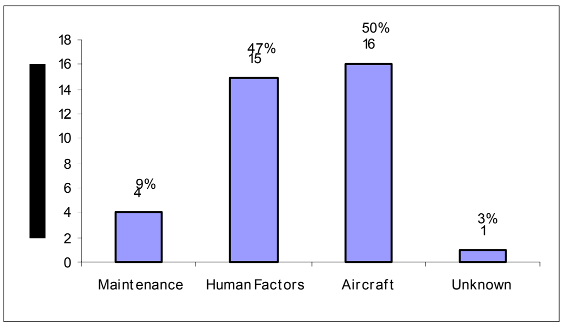



Figure 1. U.S. Army Hunter accident causal factors.

Breaking down the human factors issues further, Table 1 shows how the number and percentage of the 15 human factors-related accidents are associated with specific human factors issues. Again, percentages exceed 100% because of some accidents being classified under more than one issue.

Table 1. Breakdown of human factors issues for Hunter accidents.

| Issue | Number | Percent |
|---|---|---|
| Pilot-in-command | 1 | 7% |
| Alerts and Alarms | 2 | 13% |
| Display Design | 1 | 7% |
| External Pilot Landing Error | 7 | 47% |
| External Pilot Takeoff Error | 3 | 20% |
| Procedural Error | 3 | 20% |

By far the largest human factors issue is the difficulty experienced by EPs during landings. Forty-seven percent of the human factors-related Hunter accidents involved an error by the EP during landing. An additional 20% of the accidents involved an error by the EP during takeoff. Control difficulties are at least partially explainable by the fact that when the aircraft is approaching the EP the control inputs to maneuver the aircraft left and right are opposite what they would be when the aircraft is moving away from the EP. This cross-control problem is present for any UA operated by an external pilot via visual contact.

Besides EP control problems, other issues represented in the table include pilot-in-command issues, alerts and alarms, display design, and crew procedural error. A pilot-in-command issue is a situation where the authority of the controlling pilot is superceded by other personnel in the area, violating the principle that the pilot of the aircraft has the final decision-making authority during a flight. In contrast, alerts and alarms deal with situations where a non-normal flight condition (e.g., high engine temperature) is not conveyed effectively to the crew. Display design issues typically manifest when not all of the information required for safe flight is conveyed effectively to the crew.

Finally, the crew procedural errors referred to here involved three occasions where the crew failed to properly follow established procedures. On one occasion an improper start-up sequence led to data link interference from the backup GCS. On another occasion the crew failed to follow standard departure procedures and the UA impacted a mountain. On a third occasion an EP failed to complete control box checks prior to taking control of the UA and did not verify a box switch that was in the wrong position.

## Shadow

Unlike the Hunter, the Shadow does not use an external pilot, depending instead on a launcher for takeoffs, and an automated landing system for recovery. The landing system, called the tactical automated landing system (TALS) controls the aircraft during approach and landing, usually without intervention from the GCS pilot. A cable system, similar to the one used for the Hunter, is used to stop the aircraft after landing. Aircraft control during flight is accomplished by the GCS pilot through a computer menu interface that allows selection of altitude, heading, and airspeed. During landing, GCS personnel have no visual contact with the aircraft, nor do they have any sensor input from onboard sensors. A command to stop the aircraft engine is given by the GCS pilot, who must rely on an external observer to communicate that the plane has touched down.

The analysis of Shadow accidents shows a different pattern from that seen with the Hunter. In contrast to the Hunter, only 5 of the 24 Shadow accidents (21%) were attributed to human factors issues. Figure 2 shows the major causal factors for the Shadow accidents.

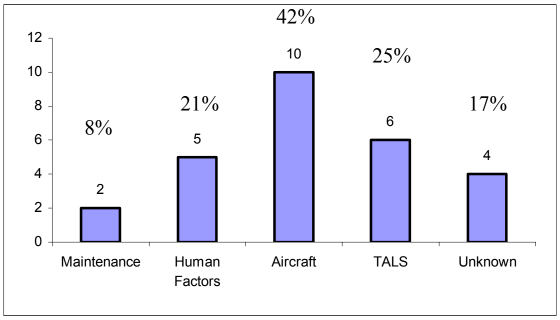



Figure 2. U.S. Army Shadow accident causal factors.

In addition to the four categories used for the Hunter accidents, an additional category was added for Shadow to include failures of the tactical automated landing system (TALS). While eliminating landing accidents potentially attributable to an EP, the use of TALS is not perfect, as shown from the data. Use of the launcher eliminated any EP takeoff errors for these aircraft.

Breaking down the human factors-related accidents, Table 2 shows the number and percentage of the 5 accidents related to specific human factors issues. As can be seen from the table, the distribution of issues is evenly divided across pilot-in-command, alerts and alarms, display design, and procedural errors.

Table 2. Breakdown of human factors issues for Shadow accidents.

| Issue | Number | Percent |
|---|---|---|
| Pilot-in-command | 2 | 40% |
| Alerts & Alarms | 2 | 40% |
| Display Design | 2 | 40% |
| Procedural error | 2 | 40% |

For both the Hunter and Shadow, at least one accident involved the transfer of control of the aircraft from one GCS to another during flight, an activity unique to UA. In the case of the Shadow, two aircraft were damaged during a single mission. The first was damaged due to a TALS failure. After the accident, the GCS crew issued a command to the damaged aircraft to kill its engine, but because of damage to the antenna the command was not received. That same GCS was then tasked with controlling a second Shadow that was on an approach. Unfortunately, after taking control of the second Shadow, the aircraft received the "engine kill" command that was still waiting for an acknowledgment from the GCS software, causing the second Shadow to crash also. This accident was classified as both a procedural error, because the crew failed to follow all checklist items prior to the transfer of control of the second aircraft, and a display design problem, because there was not a clear indication to the crew of the status of the "engine kill" command that had been issued.

## Pioneer

Like the U.S. Army's Hunter UA, the Pioneer requires an EP for takeoff and landing. After takeoff, the aircraft can be controlled from a GCS in one of three modes. In the first mode the air vehicle is operated autonomously and the autopilot uses global positioning system (GPS) preprogrammed coordinates to fly the air vehicle to each waypoint. In the second mode, the IP commands the autopilot by setting knobs (rotary position switches) to command airspeed, altitude, compass heading or roll angle, and the autopilot flies the UA. In the third mode, the IP flies the aircraft using a joystick. The Pioneer can be landed at a runway using arresting cables, but because it is a U.S. Navy/Marine operated aircraft, it is also landed on board a ship by flying into a net. There are plans for implementing an automated landing system for the Pioneer for ship-based landings.

A list of 239 Pioneer accidents was received from the Navy Safety Center. Although not providing much detail, the data did allow a general categorization of accidents into principle causal categories. Figure 3 shows the major causal factors for Pioneer accidents.

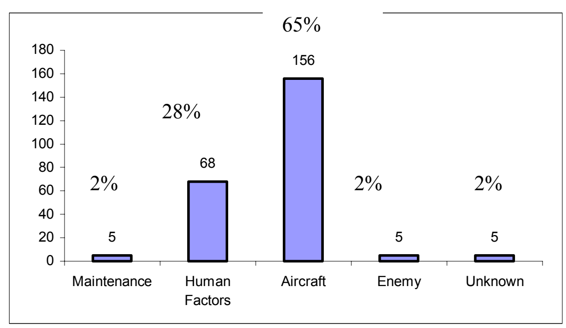



Figure 3. U.S. Navy Pioneer UA accident causal factors.

As can be seen from the figure, human factors-related issues were present in approximately 28% of the accidents. Breaking down the human factors-related accidents further, Table 3 lists the number and percentage of the 68 accidents related to specific human factors issues.

Table 3. Breakdown of human factors issues for Pioneer accidents.

| Issue | Number | Percentage |
|---|---|---|
| Aircrew Coordination | 9 | 13% |
| Landing Error | 46 | 68% |
| Take-off Error | 7 | 10% |
| Weather | 6 | 9% |

As with the U.S. Army Hunter accidents, the largest percentage of human factors accidents (68%) was associated with the difficulty experienced by the EP while landing the aircraft. An additional 10% of the accidents were associated with takeoffs, although the primary means of taking off is through the use of a launcher (from ship-based aircraft). In addition to landing and takeoff errors, two other issues seen with the Pioneer were aircrew coordination, which includes procedural and communication type errors, and weather-related accidents, which deal with pilot decision-making. Unfortunately, details regarding these accidents were not sufficient to identify issues beyond this level.

## Predator

The Predator made its first flight in June 1994. There are two Predator types, currently designated as MQ-1 and MQ-9, also called Predator and Predator B. The Predator aircraft is flown from within the GCS, similarly to a manned aircraft, using a joystick and rudder pedals and a forward-looking camera that provides the pilot with a 30-degree field of view. The camera is used for both takeoffs and landings.

The Predator accident causal factors are shown in Figure 4. As can be seen from the figure, human factors encompass a higher percentage (67%) than aircraft-related causes, unlike the other aircraft examined thus far.

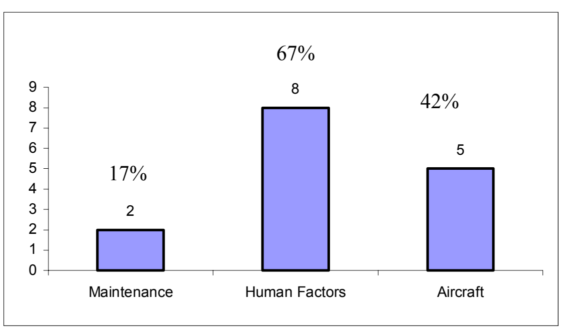



Figure 4. Air Force Predator accident causal factors.

Table 4 shows a breakdown of the human factors issues associated with Predator accidents. The majority of human factors-related problems were concerned with procedural errors on the part of the flight crew. One of these accidents involved yet another problem with a handoff of the aircraft from one GCS to another. During the handoff, the mishap crew did not accomplish all of the checklist steps in the proper order, resulting in turning off both the engine and the stability augmentation system of the aircraft. The aircraft immediately entered an uncommanded dive and crashed.

Table 4. Breakdown of human factors issues for Predator accidents.

| Issue | Number | Percentage |
|---|---|---|
| Alerts & Alarms | 1 | 13% |
| Display Design | 2 | 25% |
| Landing Error | 1 | 13% |
| Procedural Error | 6 | 75% |

A second procedural error of note occurred when the pilot accidentally activated a program that erased the internal random access memory on board the aircraft during a flight. That this was even possible to do during a flight is notable in itself and suggests the relatively ad hoc software development process occurring for these systems (Tvaryanas, 2004).

## Global Hawk

The Global Hawk, made by Northrop Grumman, is the largest and newest of the 5 military systems discussed. The first flight of the Global Hawk occurred in February 1998, and it became the first UA to cross the Pacific Ocean in April 2001 when it flew from the United States to Australia (Schaefer, 2003).

The Global Hawk is the most automated of all the systems discussed. All portions of the flight, including landing and takeoff are pre-programmed before the flight and the basic task of the crew during the flight is simply to monitor the status of the aircraft and control the payload. While this makes flying the Global Hawk very simple, the mission planning process is unwieldy and requires a great deal of time to accomplish.

Only three accident reports were available for the Global Hawk. Of these three reports, one did not provide sufficient information for classification, a second faulted a failure in a fuel nozzle, which led to an engine failure, and the third was a human factors issue centering on the complicated mission planning process. In that accident, the mishap aircraft suffered an inflight problem with temperature regulation of the avionics compartment and landed at a preprogrammed alternate airport for servicing. After landing, the aircraft was commanded to begin taxiing. Unknown to the crew, a taxi speed of 155 knots had been input into the mission plan at that particular waypoint as a result of a software bug in the automated mission planning software in use at the time. The aircraft accelerated to the point it was unable to negotiate a turn and ran off of the runway, collapsing the nose gear and causing extensive damage to the aircraft.

## Conclusions

One conclusion apparent from the data reported here is that, for most of the systems examined, electrical and mechanical reliability play as much or more of a role in the accidents as human error. Mishaps attributed at least partially to aircraft failures range from 33% (Global Hawk) to 67% (Shadow) in the data reported here.

An improvement in electromechanical reliability will probably come only through an increase in the cost of the aircraft. However, a reduction of human errors leading to accidents might not necessarily entail increased costs if suggested changes can be incorporated early in the design process. In the systems analyzed, human factors issues were present in 21% (Shadow) to 67% (Predator) of the accidents. These numbers suggest there is room for improvement if specific human factors issues can be identified and addressed.

In that regard, it is important to note that many of the human factors issues identified are very much dependent on the particular systems being flown. For example, both the Pioneer and Hunter systems have problems associated with the difficulty external pilots have in controlling the aircraft. For both of these systems, the majority of accidents due to human error can be attributed to this problem. However, the other three systems discussed do not use an EP and either use an IP (Predator) or perform landings using an automated system (Shadow and Global Hawk).

The design of the user interfaces of these systems are, for the most part, not based on previously established aviation display concepts. Part of the cause for this is that the developers of these system interfaces are not primarily aircraft manufacturers. Another reason is that these aircraft are not "flown" in the traditional sense of the word. Only one of the aircraft reviewed (Predator) has a pilot/operator interface that could be considered similar to a manned aircraft. For the other UA, control of the aircraft by the GCS pilot/operator is accomplished indirectly through the use of menu selections, dedicated knobs, or preprogrammed routes. These aircraft are not flown but "commanded." This is a paradigm shift that must be understood if appropriate decisions are to be made regarding pilot/operator qualifications, display requirements, and critical human factors issues to be addressed.

If the aircraft is commanded to begin taxiing, there should be information available regarding the intended taxi speed. If the aircraft is being handed off from one station to another, the receiving station personnel should be aware of what commands will be transmitted to the aircraft after control is established. Interface development needs to be focused around the task of the pilot/operator. For most of these aircraft, that task is one of issuing commands and verifying that those commands are accepted and followed. Understanding this task and creating the interface to support it should help to improve the usability of the interface and reduce the number of accidents for these aircraft. This is especially important as these aircraft begin to transition to the National Airspace System (NAS), conducting civilian operations in among civilian manned aircraft.

## References

Manning, S.D., Rash, C.E., LeDuc, P.A, Noback, R.K., & McKeon, J. (2004). The role of human causal factors in U.S. Army unmanned aerial vehicle accidents. U.S. Army Aeromedical Research Laboratory Report # 2004-11.

Schaefer, R. (2003). Unmanned aerial vehicle reliability study. Office of the Secretary of Defense, Washington, DC, February 2003.

Tvaryanas, A.P. (2004). USAF UAV mishap epidemiology, 1997-2003. Presented at the *Human Factors of Uninhabited Aerial Vehicles First Annual Workshop*, Phoenix, AZ, May 24-25, 2004.

# HUMAN FACTORS CONCERNS IN UAV FLIGHT

Jason S. McCarley & Christopher D. Wickens
Institute of Aviation, Aviation Human Factors Division
University of Illinois at Urbana-Champaign

Unmanned aerial vehicles have potential to serve a range of applications of civil airspace. The UAV operator's task, however, is different from and in some ways more difficult than the task of piloting a manned aircraft. Standards and regulations for unmanned flight in the national airspace must therefore pay particular attention to human factors in UAV operation. The present work discusses a number of human factors issues related to UAV flight, briefly reviews existing relevant empirical data, and suggests topics for future research.

## Introduction

System developers have proposed a wide range of government, scientific, and commercial applications for unmanned aerial vehicles (UAVs), including border and port security, homeland surveillance, scientific data collection, cross-country transport, and telecommunications services. Before these possibilities can be realized, however, FAA standards and regulations for UAV operations in the NAS must be established. Given the military's experience that accident/incident rates for UAVs are several times higher than those for manned aircraft (Williams, 2004), the import of carefully designed standards and regulations for UAV flight is clear. Human factors issues are likely to be of particular concern in establishing guidelines for safe UAV flight. As noted by Gawron (1998), UAV flight presents human factors challenges different from and beyond those of manned flight, arising primarily because the aircraft and its operator are not colocated. The goal of the current work is to identify human factors issues in UAV operations, and to review relevant studies in the existing literature. The present document provides a preliminary summary of this work.

Issues discussed below will be grouped into the categories of *Displays and Controls*; *Automation and System Failures*; and *Crew Composition, Selection and Training*. As will be clear, however, the topics presented within various categories are highly interrelated. Answers to questions about crew complement, for example, are likely to depend in part on the nature and reliability of automation provided to support UAV operators. The nature of automation required for safe UAV operation, in turn, is likely to depend in part on the quality of displays and controls provided to the UAV operator.

## Displays and Controls

One of the primary consequences of the separation between aircraft and operator is that the operator is deprived of a range of sensory cues that are available to the pilot of a manned aircraft. Rather than receiving direct sensory input from the environment in which his/her vehicle is operating, a UAV operator receives only that sensory information provided by onboard sensors via datalink. Currently, this consists primarily of visual imagery covering a restricted field-of-view. Sensory cues that are lost therefore include ambient visual information, kinesthetic/vestibular input, and sound. As compared to the pilot of a manned aircraft, thus, a UAV operator can be said perform in relative "sensory isolation" from the vehicle under his/her control. Research is necessary to identify specific ways in which this sensory isolation affects operator performance in various tasks and stages of flight, and more importantly, to explore advanced display designs which might compensate for the lack of direct sensory input from the environment.

Work by Ruff, et al (2000), Calhoun, et al (2002), and Dixon, et al (2003) has begun to address to these issues by exploring the benefits of multimodal displays to UAV operators. Ruff and colleagues examined the utility of haptic displays for alerting UAV operators to the onset of turbulence. To the pilot of a manned aircraft, turbulence is signaled by visual, auditory, and kinesthetic/haptic information. To the pilot of a UAV with a conventional display, in contrast, turbulence is indicated solely by perturbations of the camera image provided by the UAV sensors. A study by Ruff, et al, found that haptic information conveyed via the joystick control improved operator's self-rated situation awareness in a simulated UAV approach and landing task. These improvements obtained, however, only under limited circumstances (specifically, only when the turbulence occurred far from the runway; no benefits to SA were observed when turbulence occurred near the runway) and were offset by an increase in the subjective difficulty of landing. These results suggest some value of multi-modal displays as a method of compensating for sensory information denied to a UAV operator with conventional displays, but indicate that such displays may carry performance costs as well. Future research is necessary to examine the costs and benefits of multimodal displays in countering for UAV operators' sensory isolation, and to determine the optimal design of such displays.

A related point is that multimodal displays may be useful not simply as a means to compensate for the UAV operator's impoverished sensory environment, but more generally to reduce the cognitive and perceptual workload levels. Studies by Calhoun, et al (2002) and Dixon, et al (2003), for example, tested the value of tactile and auditory displays, respectively, as a method of alerting operators to system failures. Given the high visual demands of the UAV flight control task, the experimenters predicted such multimodal displays would enable better human performance than would visual displays of system status (Wickens, 2000). Consistent with this prediction, system failures in these studies were detected more quickly when signaled through tactile or auditiory displays than when indicated visually. Data from Calhoun, et al (2002) suggested that multimodal displays, by offloading of workload from the visual channel, can improve flight tracking performance. Additional research should further address the value of multimodal displays for offloading visual information processing demands. A related point is that multimodal operator controls (e.g., speech commands) may also help to distribute workload across sensory and response channels (Draper, et al, 2003; Gunn, et al, 2002), and should be explored.

An additional concern imposed by the separation between vehicle and operator is that the quality of visual sensor information presented to the UAV operator will be constrained by the bandwidth of the communications link between the vehicle and its ground control station. Data link bandwidth limits, for example, will limit the temporal resolution, spatial resolution, color capabilities and field of view of visual displays (Van Erp, 1999), and data transmission delays will delay feedback in response to operator control inputs. Research is necessary to examine the design of displays to circumvent such difficulties, and the circumstances that may dictate levels of tradeoffs between the different display aspects (e.g., when can a longer time delay be accepted if it provides higher image resolution). Research has found, not surprisingly, that a UAV operators' ability to track a target with a payload camera is impaired by low temporal update rates and long transmission delays (Van Erp & Breda, 1999). Additional research should be conducted to determine the effects of lowered spatial and/or temporal resolution and of restricted field of view on other aspects of UAV and payload sensor control (e.g., flight control during takeoff and landing, traffic detection). Of further interest is the possibility of augmented reality and/or synthetic vision systems (SVS) to supplement sensor input (Draper, et al, 2004). Studies by Van Erp & Van Breda (1999) have found that such augmented reality displays can improve the accuracy and reduce the cognitive demands of target tracking with a payload sensor, and by extension improve UAV flight control.

**Automation and System Failures**

Current UAV systems differ dramatically in the degree to which flight control is automated. In some cases the aircraft is guided manually using stick and rudder controls, with the operator receiving visual imagery from a forward looking camera mounted on the vehicle. In other cases control is partially automated, such that the operator selects the desired parameters through an interface in the ground control station. In other cases still control is fully automated, such that an autopilot maintains flight control using preprogrammed fly-to coordinates. The manner of flight control used during takeoff and landing, further, often differs from the manner of control used en route. The relative merits of each form of flight control may differ as a function of the time delays in communication between operator and UAV and the quality of visual imagery and other sensory information provided to the operator from the UAV. Research is needed to determine the circumstances (e.g., low time delay vs. high time delay, normal operations vs. conflict avoidance and/or system failure modes) under which each form of UAV control is optimal. Of particular importance will be research to determine the optimal method of UAV control during takeoff and landing, as military data indicate that a disproportionate number of the accidents for which human error is a contributing factor occur during these phases of flight (Williams, 2004).

Research will also be necessary to examine the interaction of human operators and automated systems in UAV flight. A study by Dixon & Wickens (2003) found that allocation of flight control to an autopilot freed attentional resources and improved performance on a concurrent visual target and system fault detection tasks. This effect obtained even if the autopilot was not perfectly reliable but occasionally drifted off course. The converse effect, however, did not hold; automated auditory alerts to signal the occurrence of system faults produced no benefit to flight tracking performance. The benefits of automation are also likely to depend on the level at which automation operates (Mouloua, et al, 2001; Parasuraman, et al, 2000). For example, Ruff, et al (2002) found different benefits for automation managed by consent (i.e., automation which recommends a course of action but does carry it out until the operator gives approval) and automation managed by exception (i.e., automation which carries out a recommended a course of action unless commanded otherwise by the operator) in a simulated UAV supervisory monitoring task. Research is thus needed to determine which of the UAV operator's tasks (e.g., flight control, traffic detection, system failure detection) should be automated and what levels of automation are optimal. A corollary of these recommendations is that research will be necessary to establish and optimize procedures for responding to automation or other system failures. For example, it will be important for the UAV operator and air traffic controllers to have clear expectations as to how the UAV will behave in the event that communication with the vehicle are lost.

**Crew Composition, Coordination, Selection, and Training**

A third set of human factors-related issues pertains to the composition, selection, and training of UAV flight crews. UAV flight crews for military reconnaissance missions typically comprise two operators, with one responsible for airframe control and the other for payload sensor control.

Such crew structure is merited in light of findings that the assignment of airframe and payload control to a single operator with conventional UAV displays can substantially degrades performance (Van Breda, 1995). Data also suggest, however, that appropriately designed displays and automation may help to mitigate the costs of assigning UAV and payload control to a single operator (Dixon, et al, 2003; Van Erp & Van Breda, 1999). It may even be possible for a single UAV operator to monitor and supervise multiple semi-autonomous vehicles simultaneously. Study is necessary to determine crew size and structure necessary for various categories of UAV missions in the NAS, and to explore display designs and automated aids that might reduce crew demands and potentially allow a single pilot to operate multiple UAVs simultaneously. Research is necessary on techniques to understand (Gorman, et al, 2003) and facilitate (Draper, et al, 2000) crew communications, with perhaps particular focus on inter-crew coordination during the hand off of UAV control from one team of operators to another (Williams, 2004).

Finally, study is necessary to examine standards for selecting and training UAV operators. There are currently no uniform standards across branches of the US military for UAV pilot selection; while the Air Force exclusively selects military pilots as UAV operators, Navy and Marine UAV operators are required only to have a private pilot's license, and operators of the Army's Shadow UAV generally are not rated pilots. Thus, while data from Schreiber, et al (2002), indicate significant positive transfer from manned flight experience to Predator UAV control, research is needed to determine whether such experience should be required of UAV operators. Efforts are also necessary to determine the core content of ground school training for UAV operators, and to explore flight simulation techniques for training UAV pilots (Ryder, et al, 2001).

## REFERENCES


Calhoun, G.L., Draper, M.H., Ruff, H.A., & Fontejon, J.V. (2002). Utility of a tactile display for cueing faults. Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, 2144-2148.

Dixon, S.R., Wickens, C.D. & Chang, D. (2003). Comparing quantitative model predictions to experimental data in multiple-UAV flight control. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, 104-108.

Draper, M., Calhoun, G., Ruff, H., Williamson, D., & Barry, T. (2003). Manual versus speech input for unmanned aerial vehicle control station operations. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, 109-113.

Draper, M.H., Geiselman, E.E., Lu, L.G., Roe, M.M., & Haas, M.W. (2000). Display concepts supporting crew communications of target location in unmanned air vehicles. *Proceedings of the IEA 2000/ HFES 2000 Congress*, 3.85 - 3.88.

Gawron, V.J., (1998). Human factors issues in the development, evaluation, and operation of uninhabited aerial vehicles. *AUVSI '98: Proceedings of the Association for Unmanned Vehicle Systems International*, 431-438.

Gorman, J.C., Foltz, P.W., Kiekel, P.A., Martin, M. J., & Cooke, N. J. (2003). Evaluation of latent-semantic analysis-based measures of team communications. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, 424-428.

Gunn, D.V., Nelson, W.T., Bolia, R.S., Warm, J.S., Schumsky, D.A., & Corcoran, K.J. (2002). Target acquisition with UAVs: Vigilance displays and advanced cueing interfaces. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 1541-1545.

Mouloua, M., Gilson, R., Daskarolis-Kring, E., Kring, J., & Hancock, P. (2001). Ergonomics of UAV/UCAV mission success: Considerations for data link, control, and display issues. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 144-148.

Ruff, H.A., Narayanan, S., & Draper, M.H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned aerial vehicles. *Presence, 11*, 335-351.

Ryder, J.M, Scolaro, J.A., Stokes, J.M. (2001). An instructional agent for UAV controller training. *UAVs- Sixteenth International Conference*, 3.1-3.11.

Schreiber, B.T., Lyon, D.R., Martin, E. L., & Confer, H.A. (2002). *Impact of prior flight experience on learning Predator UAV operator skills*. USAF Technical Report, AFRL-HE-AZ-TR-2002-0026.

Van Breda, L. (1995). *Operator performance in multi Maritime Unmanned Air Vehicle control* (Report TNO-TM 1995 A-76). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Van Erp, J.B.F., & Van Breda, L. (1999). *Human factors issues and advanced interface design in maritime unmanned aerial vehicles: A project overview* (Report TNO TM-99-A004). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 159-177.

Williams, K.W. (2004). *A summary of unmanned aerial aircraft accident/incident data: Human factors implications*. Technical report.

# Visibility in the Aviation Environment

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

Problems with visibility play an enormous role in a large number of fatalities in aviation accidents each year. These problems often occur in the context of proceeding visually into instrument meteorological conditions (IMC) and result in a variety of accidents both on the ground and in the air. The accidents not only occur due to visually demanding conditions but also because pilots sometimes fail to recognize conditions that make it difficult to detect other objects and/or may fail to take corrective action. The purpose of the present project is to develop research and educational materials that will help reduce accidents caused by problems of visibility in the aviation environment in the air and on the ground. Research includes analysis and quantification of the statistics of the aviation environment in the context of visibility and target detection. Further research is aimed at determining pilot performance as a function of these environmental statistics. The project will also advance the development of educational materials based on the results from the detection experiments.

## Introduction

*General*

The present report represents the first annual report for this project due to a late funding date of April 2003 and covers activity from April of 2003 until October of 2004. There are several important goals that have been accomplished during this period which will be described below.

*Purpose*

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

1) *Learning to see the target-* Visual detection is an active task rather than a passive one. Efficient search and detection requires that the observer know what to look for, that is approximately where, when, and how it will appear. Just as with the auditory system, the process of sensory encoding requires prior knowledge for optimal performance. Student pilots are often unable to understand what air traffic controllers are saying on frequency until they learn what to expect to hear. Similarly pilots must learn what to expect to see in order to acquire visual targets optimally. Additionally, the more salient the target is the easier it is to detect.

For example when an air traffic controller calls "traffic, Cessna, 2:00, 2 miles, southbound, 6000" for a pilot, the pilot must know first where to look. The ability to judge azimuth is usually assumed as most pilots would be familiar with "clock" directions and particularly since the information is given essentially with an angular measure that does not change with distance to the target. However elevation is not as well learned because few pilots have an intuitive feel for how high or low traffic should be given the relative altitudes of the two aircraft and the distance. In this case the pilot must determine how much of an angle to look up or down at from relative altitude and distance information. Indeed most pilots even find it difficult to determine whether or not objects such as clouds or mountains are at the same altitude as the aircraft.

Pilots must also learn what to expect to see so the pilot must be able to predict the approximate shape and size of the target aircraft. The shape can only be inferred from relative direction of travel. This has to be computed from what is known about the relative directions of the aircraft. The size must also be computed from the relative distance of the aircraft and what is known about the size of that target. In the above example "Cessna, 2 miles southbound" is the information given, so that a pilot must calculate what the target airplane should look like from this information and what is known about Cessnas and the pilot's own direction of travel. This is a complex task that requires experience to perform well.

The parameters described above are all easily calculated from known relationships. Training is required however for pilots to perform quickly and automatically.

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

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

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

In order to train pilots to judge conditions under which detection may be difficult we must first have a way to characterize the background. We must then model detection on different backgrounds composed of images from the aviation environment. We have investigated a leading model used for detection and have begun to apply the model to various images and test the model psychophysically employing detection experiments.

The results from these detection experiments should provide verification of the model of detection and evaluation of any real aviation background. This knowledge will allow us to educate pilots on recognition of dangerous conditions for detection.

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

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

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

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

*3) taking proper action*

The educational materials for the present project will be focused on training pilots to recognize demanding visual conditions. Future experiments will address issues surrounding failure of pilots to take action once difficult visual conditions are encountered and recognized (see e.g. O'Hare and Owen, 2002).

## Accomplishments and Results

*Simulator*

We have now completed the construction of simulator system with 180 deg of "outside" visual display (see fig. 1 below). This system still needs to be programmed to conduct detection and weather recognition experiments.

Fig. 1 Simulator for detection experiments.

*Aviation Images*

We have collected high quality digital images from the aviation environment over a large portion of the mainland U.S. and around the greater Anchorage area in Alaska. Many of these images have already been analyzed using sparse coding algorithms.

We have found that the basis functions "learned" by the sparse coding algorithm are different than those learned from land-based environment images. Applications of land-learned basis functions to the aviation images suggest that cortical visual development based on the terrestrial environment may not be optimal for the aviation environment. The results from this study have recently been presented at the annual Fall Vision Meeting of the Optical Society of America, in Rochester (Mizokomi and Crognale; 2004; See fig. 2 and attached poster in Powerpoint format).

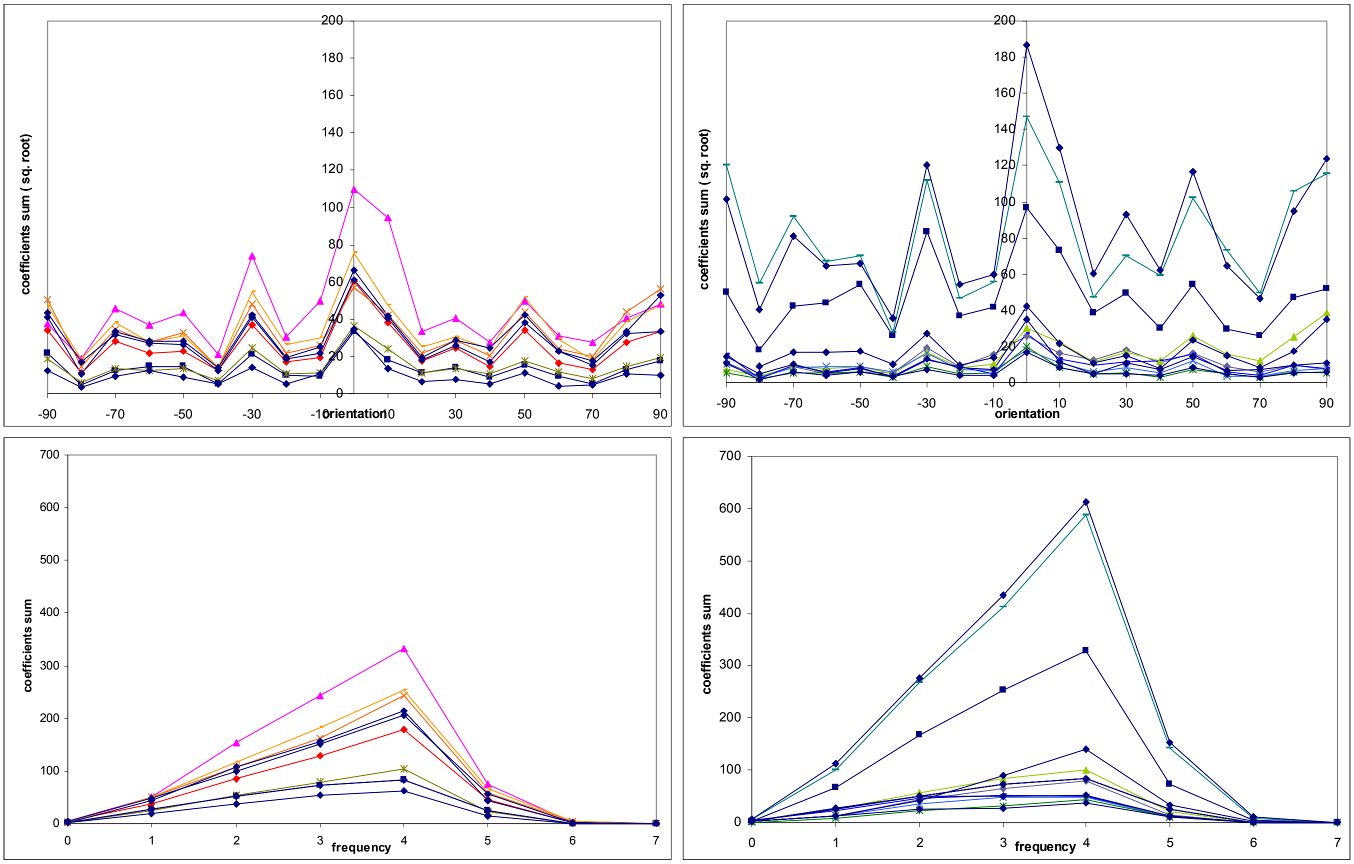



Figure 2. The weighting coefficients for orientation (upper) and spatial frequency (lower from the terrestrial learned basis functions for terrestrial images (left) and aerial images (right).

We have also started analyzing visual images from the aviation environment in terms of a visual detection model proposed by Ahumada (Ahumada, 1996;Ahumada and Beard, 1997;1998; Rohaly et al., 1997). This model has been applied to visual data sets in the terrestrial environment and well predicts detection under many conditions. The model estimates how well the detection of objects will be impaired by the background. It accomplishes this with some simple filtering algorithms that compute the contrast masking energy of the background. The model produces a measure of sensitivity (d') that should predict relative behavioral detection thresholds. Thus different aviation environments can be measured and predictions made about how difficult these environments are for detection relative to one another.

The next phase of this study will be to test the predictions of the model in behavioral detection studies both on simplified computer simulations and more advanced tasks in the flight simulator that include distractions and variables from the flying tasks.

*Learning to see*

As a preliminary step towards training pilots to see, we have developed a simple reference card for use in the cockpit (see appendix). This card illustrates the apparent sizes of typical small airplanes (e.g. Cessna 172) and airliners (e.g. Airbus A-320) at different distances from 2 miles to ½ mile. This card can be used by the pilot to estimate the approximate size of a known but undetected target. It is hoped that this aid will help improve target detection and would be especially useful for low time pilots and during private pilot training.

We have also begun to develop a preliminary version of the final training product, an interactive program that will educate and train pilots in the issues of visibility.

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

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

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

**References**


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

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

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

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

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

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

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

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

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

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

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

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

# Aid for Judging the Apparent Size of Aircraft

Print this sheet, cut out the 3 X 5 reference card, and fold it in half on the dotted line. Measure the calibration mark on the bottom and see if it measures one inch. View the card from 18 inches if the line measures one inch. If not, mutiply the standard viewing distance (18") by the length of the line. That should be your correct viewing distance for the card (example: line length = 0.8 inches. 18 X 0.8 = 14.4 “; In this case your viewing distance would be 14.4 inches).

**Know what to look for!**

Approximate appearance of an airliner and a small single engine aircraft

(view at a distance of 18")

Front

**2 Miles** **1 Mile** **1/2 Mile**

Fold Here

To quickly detect traffic, pilots should know the apparent size of aircraft at various distances. This card can be used as a reference in the cockpit and to help pilots learn target size.

To view the card, <u>hold it at a distance of 18 inches</u> from the eyes. The sizes of the images approximate those produced by actual aircraft (a Cessna 172, and an Airbus 320). These aircraft were chosen as examples of small aircraft and airliners.

**Note that the actual appearance and visibility of real aircraft will vary with color, weather, direction of travel, type of aircraft and other factors.**

Developed by Dr. Michael Crognale; Send any questions or comments regarding this aid to: Dr. Crognale (mikro@unr.edu); or The Federal Aviation Administration, General Aviation and Commercial Division; (AFS-800), Room 835, 800 Independence Avenue, S.W., Washington, DC 20591,Phone: 202-267-8212; or General Aviation Human Factors Program Manager (william.krebs@faa.gov).

Back

Cut Out

1 inch

Calibration

# The effect of terrain-depicting primary-flight-display backgrounds and guidance cues on pilot recoveries from unknown attitudes

*Dennis B. Beringer and Jerry D. Ball*
*The FAA Civil Aerospace Medical Institute*
*Oklahoma City*

*Kelly Brennan & Sitafa Taite*
*O.U. School of Industrial Engineering*
*Norman, OK*

*A study was conducted to evaluate the effects of primary flight display (PFD) terrain depictions on pilots' performance of recoveries from unknown attitudes. Forty pilots participated in the study, each group of eight using a different display format. The five conditions consisted of combinations of terrain depiction (none, full-color terrain, brown terrain) and guidance indications (pitch and roll arrows). Participants flew baseline trials in the Advanced General Aviation Research Simulator using a common electronic attitude indicator and then performed recoveries from unknown attitudes (UARs) using one of the PFD formats. Performance measures included initial response time, total recovery time, primary reversals, and secondary reversals. No significant effects of the primary independent variables were found on any of the performance measures. Posttest interviews indicated the participants preferred the directional-arrow indicators and had no preference for or against the presence of terrain depictions during UARs, focusing primarily on the zero-pitch line as a reference. It was concluded that the specific terrain representations examined did not pose a hazard to the identification of and recovery from unknown attitudes as long as a zero-pitch line of sufficient contrast (white with black borders) to all backgrounds was present.*

## BACKGROUND

Electronic Flight Instrumentation Systems (EFIS) are becoming more available daily, and a major component of this type of system is the Primary Flight Display (PFD). While PFDs initially depicted attitude and flight-guidance information, they evolved to include forward-looking perspective-views of both guidance information (Beringer, 2000) and of the outside world (Wickens, Haskell, and Hart, 1989; Alter, Barrows, Jennings, and Powell, 2000), often generated from terrain databases. This type of display is presently appearing in systems submitted for certification in general aviation (GA) aircraft, and a number of questions have been raised regarding the effects of various design features on different aspects of pilot performance. In lieu of empirical data on the effects of manipulations of specific design parameters, certifiers have had to rely upon general guidelines and often to adopt very conservative criteria for the certification and use of these particular displays.

Some data have become available, relevant to the GA environment, that may be useful for determining what the allowable range of variation in design parameters can be. The parameters that seem to be of present interest include the following: size of the display, angular representation of the outside world (field of view), display resolution, terrain-feature resolution, use of color, style of terrain representation, definition of display clutter, and effects of the above on the performance of both routine and non-routine flight tasks.

A series of studies were performed at the NASA Langely Research Center examining the use of various terrain representations and pilot preferences for various fields of view and styles of depiction (Prinzel, Hughes, Arthur, et. al., 2003; Arthur, Prinzel, Kramer, Parrish, and Bailey, 2004). Some agreement was found with previous studies concerning preference for field of view (30 degrees), and some assessment was made of pilot navigation performance and some basic precision maneuvers, concluding that fewer errors were committed and terrain awareness was enhanced with the displays. One issue that was not addressed, however, was the recovery from unknown or unusual attitudes. This specific concern was addressed in one certification process by requiring that the terrain depiction be removed from the PFD when the aircraft exceeded certain pitch or roll criteria because of a concern that the presence of the terrain might cause confusion or somehow interfere with a successful recovery. However, there were no empirical data to indicate what role, positive or negative, the terrain depiction might play in the recoveries.

Thus, a study was conducted to examine how various forms of terrain depiction might either impede or enhance recoveries from unknown attitudes, including the display content (type of terrain; flat, mountainous) at the time of the recovery as well as the possible ameliorating effect of providing recovery guidance arrows (Gershzohn, 2001). Questions of specific interest were if pilots would recover to the terrain horizon rather than the zero-pitch line if the two were different, if this behavior (if observed) could be ameliorated by positive guidance cues, and if the coloration of the terrain presentation had an effect upon performance.

## METHOD

*Experimental Display Formats*

Forty pilots participated in the study, each group of eight using a different display format. The five conditions consisted of combinations of terrain depiction (none, full-color terrain, brown terrain) and guidance indications (pitch and roll arrows). The no-terrain display consisted of a traditional attitude indicator (blue sky, brown ground) with airspeed, altitude and vertical speed presented in tape format along the left and right edges of the display with a compass card at the bottom of the display.

The second display was identical to the first, but had guidance arrows for pitch and roll recovery. Pitch arrows were linear (Figure 1) and appeared when the aircraft attitude was greater than 13 degrees up or down and disappeared when the aircraft was within 5 degrees of zero pitch, pointing from the aircraft symbol to the horizon. Roll arrows (Figure 2) were curvilinear (arc form) and appeared when the aircraft exceeded 25 degrees of bank and disappeared when the aircraft was within 10 degrees of zero bank, pointing from the plane of the wings to the horizon line. For pitch-down attitudes, the roll-command arrow took precedence over the pitch-command arrow. For pitch-up attitudes, the priority was reversed.

The third display was similar to the first except that the brown portion of the display was replaced with photo-realistic (full-color) terrain (this terrain format is shown in both Figures 1 and 2). The terrain was generated using variable-sized polygons which had photo-realistic texture applied to them to create the out-the-window scene. This is somewhat different from some other terrain-creation methods seen on other terrain-depicting displays where equal-sized polygons or even squares are used to create the terrain skin and a more generic type of texture is applied.

Figure 1. PFD with pitch-recovery arrow shown.

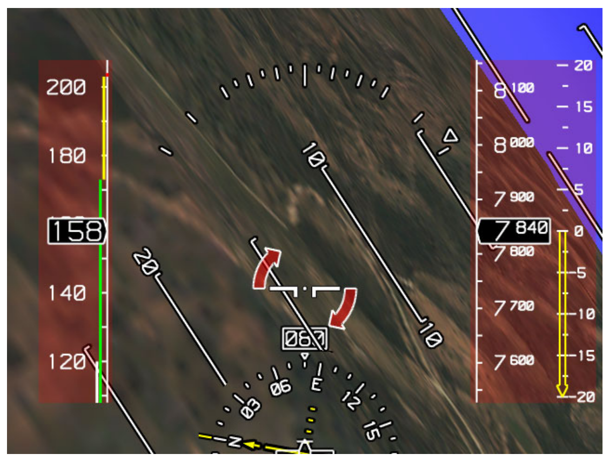


Figure 2. PFD with roll-recovery arrows shown.

The fourth display was the same as the third display, but it included the guidance arrows. The final display was similar to the first display, but the "ground" or brown portion of the display was replaced with brown (polygon-based) terrain imagery. The variable-sized polygon structure imparted more apparent texture to this uniform-brown depiction then one sees in brown-only depictions that use a uniformly sized polygon or square as the basis for terrain-contour construction. Figures 3 and 4 show similar views of a mountain in the full-color mode (Figure 3) and the brown-only mode (Figure 4) for comparison.

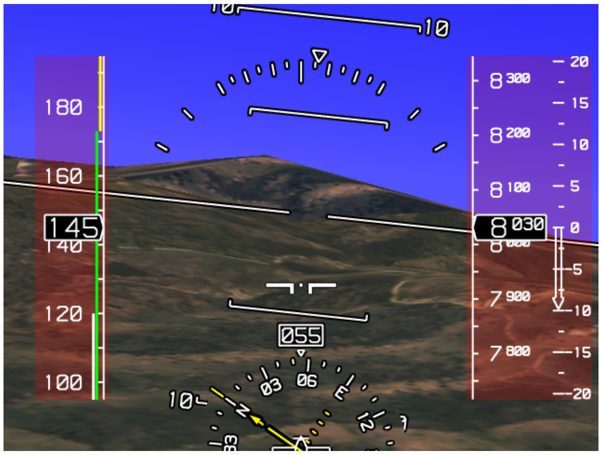


Figure 3.  PFD full-color terrain depiction with mountain in view.

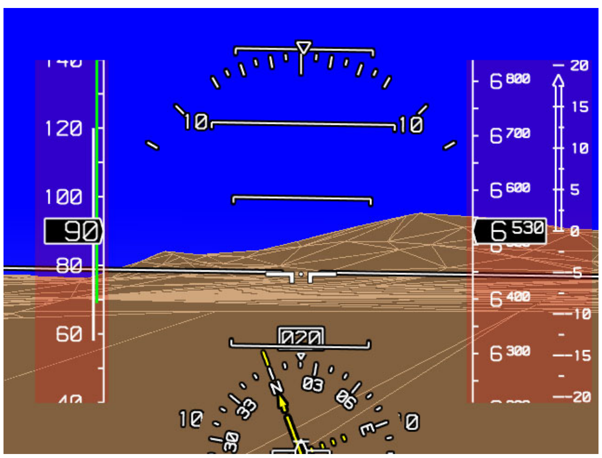


Figure 4.  PFD brown-only terrain depiction with mountain in view.

*Experimental Design*

The design was a two-factor crossed, with terrain background (full-color; present or absent) and guidance arrows (present or absent) as the *independent variables*.  The supplemental condition, brown-only terrain, was added after contribution of guidance arrows had been assessed.  Dependent variables included initial response time (IRT; time to first control input), total recovery time (TRT), primary control-input reversals, and secondary control-input reversals.

Two *sampling variables* were added to obtain more representative data from across a wider range of display indications. *Terrain depiction at roll-out* was planned using lead headings based upon expected roll-out times (obtained in pretest) and presented terrain either (1) higher than the zero-pitch reference line (mountainous background) or (2) terrain lower than the zero-pitch reference line (level terrain).  *Attitude at recovery onset* was also varied so that trials included combinations of pitch (+20, 0, and –15 degrees) and bank (60 degrees left, 0, 60 degrees right) excepting, of course, the zero-zero condition.

Three supplemental trials were also added for approximately the last 7 pilots in each group.  For each of these, a 40-degree FOV trial was added, followed by an inverted-recovery trial (by sponsor request), and finally a near-mountains trial where Sandia Peak filled the display up to the 10-degree pitch-up line when the aircraft was approaching at approximately 8000 feet MSL (the terrain horizon was significantly above the zero-pitch line).

*Equipment and participants*

Data were collected using the Advanced General Aviation Research Simulator (AGARS) in the CAMI Human Factors Research Laboratory.  The simulator was configured to represent a Piper Malibu, and the participants all flew in the left seat.  The PFD was represented on a flat-panel high-resolution LCD mounted on the instrument panel directly in front of the participant.  The PFD was presented at the size of an approximately 7-inch diagonal measurement within a larger hardware-display area, and the image showed approximately 30 horizontal degrees of the outside world.  The display layout was similar in many respects to one already certified for GA use.  The experimenter-pilot (EP) flew from the right seat with a repeater display of the PFD mounted atop the glare shield.  The out-the-window view represented a hard-IFR situation with no environmental visual cues visible in the uniformly gray fields.  Performance data were recorded digitally with supplemental audio and visual data recorded on DVD from two video sources (cockpit wide view and PFD inset) and all audio sources (participant, EP, data-collection experimenter).

Participants were 40 general aviation pilots recruited from the local community, 8 assigned to each of the five display conditions.  Age and overall flight hours were balanced across groups as participants entered the experiment (not assigned a priori from a known sample).  All were at a minimum certified as Private Pilot, while many were instrument rated and a number were flight instructors.  Each group had a similar distribution of pilot categories represented.

*Procedures/tasks*

After completing the informed consent form and filling out a brief pilot-experience questionnaire, participants were briefed concerning the display they would be using and instructed that recoveries would be from unknown attitudes. Their task was to recover to a zero-pitch zero-bank attitude regardless of altitude or airspeed, as the EP would configure the aircraft such that performance was usually within the operating envelope (primary interest was in participant ability to interpret the display and determine when a level attitude had been restored). They were then ushered into the AGARS where they were further familiarized with the display and with the simulator. They then donned a headset and a visor so that direct vision of the display would be obscured when they were in the head-down preparatory position for the recovery.

Each pilot then took off from Albuquerque (ABQ) and climbed out to the north into IFR conditions. All pilots performed 8 warm-up (baseline) recovery maneuvers, using the basic electronic attitude-direction indicator (EADI) on the PFD, to familiarize them with the performance of the AGARS and with the dynamic functioning of the PFD. Each trial began with the participant being instructed to put their head down and take their hands off of the controls. The EP then placed the simulator into the required attitude and heading for that trial, using predetermined airspeed, altitude, and heading criteria that had been rehearsed (the same EP performed all unknown-attitude entries for all participants). The EP gave a preparatory "Ready" about two seconds before handing over the controls, "and" about one second before, and "Go!" at the transfer of controls to the participant. After completing the warm-up trial, the participant flew the simulator back to ABQ and performed a full-stop landing. At this time the display format was changed and the procedure repeated.

Experimental trials consisted of 16 recovery maneuvers, defined by combinations of the sampling variables described earlier, using the PFD that was assigned to the participant. Two different orders of the combinations of sampling variables (attitude at onset and terrain seen at roll-out) were used and balanced across the groups. Accordingly, half of the headings were selected to end the recovery facing mountainous terrain higher than the aircraft altitude and half were selected to end the recovery facing terrain lower than aircraft attitude. Pilot recovery times and initial response times were recorded for each trial. A recovery was considered completed when the aircraft reached ±2.5 degrees of pitch and ±5.0 degrees of bank and was able to maintain those values for 3 seconds, although trials were generally allowed to continue for a few seconds after these criteria had been reached to guarantee stability in the recovery.

The supplemental trials were added to the end of the session. The EP flew the simulator to a designated altitude and starting point near the Sandia Mountains and one recovery was conducted where mountainous terrain occupied a significant portion of the display and the terrain horizon was 10 degrees higher than the zero-pitch line. This was followed by recovery from an inverted attitude with the nose slightly above the horizon and a bank angle of approximately 165 degrees. A final trail was flown with the display FOV changed from 30 to 40 degrees. The participant then flew the simulator back to ABQ for a full-stop landing. Participants completed a posttest set of questionnaires regarding their subjective assessment of the displays (one was also administered after the warm-up trials), went through a posttest interview, and provided both solicited and unsolicited responses/opinions.

## RESULTS

*Performance Variables*

*Recovery times*. Analysis of recovery times for the baseline trials showed that the groups initially differed in their performance, but were performing equivalently (no significant differences) by the last two trials. This finding suggests that all groups had attained a roughly equal level of performance prior to entering the experimental trials.

Multivariate Analysis of Variance indicated there were no significant differences between the display configurations for either of the response-time variables. Pitch-roll TRTs averaged around 10 seconds, whereas roll-only recoveries averaged about 8.5 seconds. Pitch-only recoveries averaged approximately 8.6 to 9.0 seconds. Univariate analyses were conducted to determine if type of maneuver resulted in any significant differences between display types. Again, no significant differences were found between displays and type of maneuver.

*Control reversals.* Examination of control reversals, defined as movements in the opposite direction of that required for the recovery, indicated that were only three clearly identifiable primary control reversals in the nearly 800 trials. There were no secondary reversals (initial response in correct direction; subsequent control movement in opposite to input required). Recovery times for the three reversals were not notably different from those of other trials. Thus, reversals did not appear to be a factor regardless of the format of display used.

*Supplemental trials*. Analyses were conducted for performance variables on each of the three supplemental trials. No significant differences were found for the 40-degree FOV trials, the inverted trials, or the near-mountains trials. Only one of the participants showed any indication of holding the nose of the aircraft above the zero-pitch line in the near-mountain trial rather than completing the recovery.

### *Questionnaires and Posttest Interviews*

Pilots indicated, when interviewed, that they were focusing their attention on the zero-pitch line, which was relatively prominent, and did not regard the terrain depictions, when present, as significant contributors to their recovery task. The directional-guidance arrows produced a positive qualitative response from the participants. Participants also expressed a relatively uniform preference for the terrain-depicting displays in general. A few individuals expressed a preference for the 40-degree FOV, stating that it allowed them to "see more." The one individual who had kept the nose of the simulator slightly higher than zero pitch for the near-mountain trial clarified, in the posttest interview, he had been concerned about the mountain and had kept the nose a little high in preparation for a possible climb over the mountain, having no indeterminacy about the zero-pitch line location.

## SUMMARY AND CONCLUSIONS

It appears, for this specific task, that the presence of a zero-pitch line of sufficient contrast (white with black borders) to all backgrounds allows pilots to adequately perform recoveries from unknown attitudes despite the specific format of perspective terrain display used in this experiment. It also appears that the directional-guidance arrows, despite being positively received by the participants and having been demonstrated to be useful in a previous experiment, did not have an appreciable effect on recovery times. The frequency of occurrence of reversals was too low to allow any conclusion to be drawn about the possible effectiveness of guidance arrows in that regard.

Given the previous findings indicating enhanced terrain awareness attributable to terrain depictions combined with the lack of detrimental effects found in this study relative to recoveries from unknown attitudes, there would appear to be fewer significant obstacles to the implementation of this type of PFD for GA use. Caveats to be observed, however, would be that (1) similarly constructed terrain depictions are used, the zero-pitch line is clearly differentiable from the terrain and sky depictions regardless of the type of background and (3) that the direction of off-display pitch-line locations are clearly indicated.

## ACKNOWLEDGMENTS

The authors would like to thank Howard Harris for contacting and scheduling the pilot participants in the study and Barry Runnels for his assistance with the operation of the AGARS. The OU authors participated in the data collection and analyses as part of their senior Capstone project in Industrial Engineering. Dr. Beringer served as their supervisor.

## REFERENCES

Alter, K. W., Barrows, A. K., Jennings, C. W., and Powell, J. D. (2000). 3-D perspective primary flight displays for aircraft. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society*, 3-29 – 3-32.

Arthur, J.J., III, Prinzel, L.J., III, Kramer, L.J., Parrish, R.V., and Bailey, R.E. (2004). Flight simulator evaluation of synthetic vision display concepts to prevent controlled flight into terrain (CFIT). Springfield, VA: NTIS, *NASA/TP-2004-213008*.

Beringer, D. B. (2000). Development of highway-in-the-sky displays for flight-path guidance: History, performance results, guidelines. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society*, 3-21 – 3-24.

Gershzohn, G., (2001). Unusual Attitude Recovery Using the Roll Arrow. *Society of* Automotive Engineers, Inc., 641-645.

Prinzel, L.J.,III, Hughes, M. F., Arthur, J.J., III, Kramer, L.J., Glaab, L.J., Bailey, R.E., Parrish, R.V., and Uenking, M.D. (2003). Synthetic vision CFIT experiments for GA and comercial aircraft: a picture is worth a thousand lives. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*.

Wickens, C. D., Haskell, II, & Harte, K. (1989). Perspective format for a primary flight display. *Tech. Rpt. ARL-89-2/Boeing-89-1*. Savoy, IL: University of Illinois, Aviation Research Lab.