2005 International Journal of Aviation Research and Development Volume 5, Number 2 A publication of the FAA Academy

International Journal of Applied Aviation Studies

A Publication of the FAA Academy Oklahoma City, Oklahoma

Volume 5, Number 2, 2005
Government Printing Office

REVIEW PROCESS

The Federal Aviation Administration Academy provides traceability and oversight for each step of the International Journal of Applied Aviation Studies (IJAAS). IJAAS is a peer-reviewed publication, enlisting the support of an international panel of consulting editors. Each consulting editor was chosen for his or her expertise in one or more areas of interest in aviation. Using the blind-review process, three or more consulting editors are selected to appraise each article, judging whether or not it meets the requirements of this publication. In addition to an overall appraisal, a Likert scale is used to measure attitudes regarding individual segments of each article. Articles that are accepted are those that were approved by a majority of judges. Articles that do not meet IJAAS requirements for publication are released back to their author or authors.

IJAAS is printed by the Government Printing Office. Individuals wishing to obtain a copy of the journal may contact Kay Chisholm by email at kay.chisholm@faa.gov, or by telephone at (405) 954-3264, or by writing to the following address:

International Journal of Applied Aviation Studies
Kay Chisholm
AMA-530
PO Box 25082
Oklahoma City, OK 73125

International Journal of Applied Aviation Studies
Volume 5, Number 2
ISSN Number: 1546-3214

1st Printing December 2005

POLICY AND DISCLAIMERS

Policy Statement: The Federal Aviation Administration (FAA) Academy strongly supports academic freedom and a researcher's right to publish; therefore, the Federal Aviation Administration Academy as an institution does not endorse the viewpoint or guarantee the technical correctness of any of the articles in this journal.

Disclaimer of Liability: With respect to articles available in this journal, neither the United States Government nor the Federal Aviation Administration Academy nor any of their employees, makes any warranty, express or implied, including the warranties of merchantability and fitness for a particular purpose, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Disclaimer of Endorsement: Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the United States Government or the Federal Aviation Administration Academy. The views and opinions of authors expressed herein do not state or reflect those of the United States Government or the Federal Aviation Administration, and shall not be used for advertising or product endorsement purposes.

John Cain	Florida Technology Institute, School of Aeronautics, Melbourne, FL
David Carmichael	Oklahoma City University, Oklahoma City, OK
Stephen M. Casner	NASA Ames Research Center, Moffett Field, CA
David Conway	Southeastern Oklahoma State University, Durant, OK
Janet Cope	Washington Consulting Group, Prescott, AZ
Brian G. Dillman	Purdue University, West Lafayette, IN
Key Dismukes	NASA Ames Research Center, Moffett Field, CA
Graham Elliott	FAA Academy, Oklahoma City, OK
Richard Fanjoy	Purdue University, West Lafayette, IN
Douglas R. Farrow	Federal Aviation Administration, Washington, DC
Kenneth H. Funk II	Oregon State University, Corvallis, OR
Trent Gabert	University of Oklahoma, Norman, OK
A. F. Grandt	Purdue University, West Lafayette, IN
Mavis Green	Harvard University, Cambridge, MA
Carla A. Hackworth	Civil Aerospace Medical Institute, Oklahoma City, OK
Jeffrey T. Hansberger	George Mason University, Fairfax, VA
Samuel B. Hendrix III	FAA Academy, Oklahoma City, OK
Judy Holcomb	FAA Academy, Oklahoma City, OK
Todd P. Hubbard	Oklahoma State University, Stillwater, OK
Michael Hynes	Hynes Aviation Services, Frederick, OK
Sig Illing	FAA Academy, Oklahoma City, OK
Merrill R. Karp	Arizona State University, Mesa, AZ
Raymond E. King	Civil Aerospace Medical Institute, Oklahoma City, OK
William Knecht	Civil Aerospace Medical Institute, Oklahoma City, OK
Jefferson M. Koonce	IFR Flight Training School, Georgetown, TX
Mary N. Kutz	Oklahoma State University, Stillwater, OK
Florian Jentsch	University of Central Florida, Orlando, FL
Victor LaSaxon	The Boeing Company, Midwest City, OK
Chien-tsung Lu	Central Missouri State University, Warrensburg, MO
Rebecca Lutte	University of Nebraska at Omaha, Omaha, NE
Beth Lyall	Research Integrations, Inc., Tempe, AZ
Richard L. Mangrum	Kent State University, Kent, OH
Carol Manning	Civil Aerospace Medical Institute, Oklahoma City, OK
Harry Marks	University of Oklahoma, Norman, OK
Steven Marks	Oklahoma State University, Stillwater, OK
Royce Ann Martin	Bowling Green State University, Bowling Green, OH
Celeste Mayer	North Carolina State University, Raleigh, NC
H. C. McClure	MAC Air Consulting, Queen Creek, AZ
Jesse McMullen	FAA Academy, Oklahoma City, OK
Marjo Mitsutomi	University of Redlands, Redlands, CA
Richard H. Mogford	NASA Ames Research Center, Moffett Field, CA
John C. Morey	Dynamics Research Corporation, Andover, MA
Kathleen Mosier	San Francisco State University, San Francisco, CA
Gary J. Northam	Parks College, St. Louis University, St. Louis, MO
Dale B. Oderman	Purdue University, West Lafayette, IN
Manoj Patankar	Parks College, St. Louis University, St. Louis, MO
Mike Pearce	American Airlines
Donald A. Petrin	Purdue University, West Lafayette, IN
Julia Pounds	Civil Aerospace Medical Institute, Oklahoma City, OK
Stephen M. Quilty	Bowling Green State University, Bowling Green, OH

PHILOSOPHY STATEMENT

Cornelius Lanczos, a mathematician working in the field of applied analysis, expressed the history of mathematics in three phases:

1) A given physical situation is translated into the realm of numbers,
2) By purely formal operations with these numbers certain mathematical results are obtained, [and]
3) These results are translated back into the world of physical reality (1988, p. 1). [1]

Formal papers, in subjects related to aviation, roughly follow the same course. However, there appears to be a weakness in aviation research, that being the omission of the third phase.

It is not good enough that conclusions are drawn, if those conclusions fail to improve the system observed. Clearly, the observed have a say in implementing the conclusions of research, but their failure to implement the conclusions drawn by the researcher may be more indicative of a lack of understanding than a lack of desire. Researchers tend to peer into complex systems as through a soda straw, forming formal opinions on the finite without understanding the complete system. Industry, ever mindful of the complete system, may find research irrelevant, because it makes much to do about nothing.

The editorial staff, to include those listed as consulting editors, is committed to the improvement of all individuals within the aviation community. We seek to enhance existing systems bearing in mind that small improvements must not upset the delicate balance between too little and too much help. We also seek to promote safety, not by lip service, but by demonstration in how we execute our studies and how we report our findings.

We feel that the best way to translate results back to the physical world is to incorporate the viewpoints of people around the globe. Without the influence of a worldwide community, we deny the significance of diversity, and ignore the perspectives of gifted scientists from different countries. It is our hope that each reader will feel the same.

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

EDITOR'S NOTES

Formal Papers

Our leading article by Hoover and Russ-Eft studies whether single-pilot multi-tasking skills can be trained. Twenty-seven pilots were randomly assigned to an experimental or control group. All pilots flew a pretest and a posttest. In the time between the two tests, pilots in the experimental group participated in a concurrent task management training course and the control group did not. The posttest results supported training multi-tasking skills in single-pilot operations.

The Hackworth, Cruz, Goldman, Jack, King, and Twohig article evaluates the findings of the FAA's 2003 Employee Attitude Survey (EAS). The 2003 EAS contained 129 items organized into three major sections: (1) Indicators of Satisfaction, (2) Management and Work Environment, and (3) Respondent Demographics. A corporate plan addressing several areas identified as needing improvement on the 2003 EAS is discussed.

To understand travelers' behavior, Bowen, Metz, and Headley used the Airline Survey for their study. The study explores the perceptions of frequent travelers, what factors go into selecting a specific airline, consumer complaints, and the impact on airline travel from the terrorist attacks of September 11, 2001. The survey results provide an indication of consumer behavior and response to the changing air transportation environment.

Since personal computer-based aviation training devices (PCATDs) have been shown to support primary flight training for fixed-wing aircraft, Johnson and Stewart investigated their value for helicopter flight training. Sixteen aviators representing both highly experienced and student helicopter pilots evaluated the ability of a commercial PCATD on the seventy-one flight tasks comprising the IERW Common Core curriculum. The PCATD was judged as best supporting Instrument Flight Training, especially tasks involving radio navigation. Tasks requiring hovering were judged as less well supported.

Casner's study addresses the transfer of skills from small technically advanced aircraft to the more sophisticated equipment found in modern jet transports. Pilots, who received training in small technically advanced aircraft, successfully completed 83% of all procedures in the jet transport, while pilots in the control group achieved an average success rate of 54%. The results support transfer of learning between the two types of equipment, and support the use of small technically advanced aircraft to train pilots who will later transition to the commercial jets.

The crash of Korean Air Flight 801 highlighted a need for research on the pilot-monitored approach. Wochinger and Boehm-Davis present the results of an initial research effort. The research used the internet to survey professional pilots about the pilot-monitored approach, and its safety and utility. The analysis focused on responses from 205 pilots. Though pilots view the pilot-monitored approach as safe, the findings also showed that there is much uncertainty about very basic aspects of the pilot-monitored approach.

Air traffic controller trainees and science students are alike in that space-time concepts and reasoning ability of facts are equal components of course curricula. This is why Baharestani's study of the differential effects of three different instructional treatments: Computer Text Instruction (CTI), Computer Text-Graphic Instruction (CTGI), and Computer-Based Instruction (CBI) is important. The results indicated that students who used the CBI lesson did significantly better than students using the CTGI or CTI lessons did.

The goal of Bustamante, Fallon, Bliss, Bailey, and Anderson's study was to examine pilots' workload, situation awareness (SA), and trust in weather systems during critical weather events as a function of time pressure, role assignment, pilots' rank, weather display, and weather system. Results showed that airlines should consider a change in role assignment philosophy, encourage pilots to make deviation decisions around weather sooner, provide pilots with more weather information using the NEXRAD system, and should explore improvements so that NEXRAD information can be presented in real time.

Developmental Papers

The traditional step-by-step model of decision making has failed to serve pilots. Bertrand's article focuses on the research that examines real world decision making by aviators and others in similarly placed positions. It examines two models of decision making that the author believes are much better suited to aviation and comments on their potential for wider application in the profession.

Book Reviews

Have we been focusing too much on teaching pilots how to fly and not enough on preparing pilots to embrace all aspects of aviation? Hubbard poses this question in the review of "Aviation Education and Training" by Irene Henley (editor). Eleven authors address issues relating to collegiate aviation programs and aviation training.

KC

International Journal of Applied Aviation Studies

Volume 5, Number 2 — 2005

International Journal of Applied Aviation Studies, 5, Number 2

Formal Papers

Effect of Concurrent Task Management Training on Single Pilot Task Prioritization Performance

Amy L. Hoover
Flight Technology Department
Central Washington University
400 E University Way
Ellensburg, WA 98926
hoovera@cwu.edu

and

Darlene F. Russ-Eft
College of Education
411 Education Hall
Oregon State University
Corvallis, OR 97331
darlene.russeft@oregonstate.edu

Abstract

This study tested whether single-pilot multi-tasking skills can be trained. Twenty-seven pilots enrolled in a university flight technology program were randomly assigned to an experimental or a control group. Pilots flew pretest and posttest simulated flights on a Frasca 141 FTD and 20 task prioritization challenges were embedded within each flight scenario. Pretest task prioritization performance of the two groups was similar. During two weeks between pretest and posttest the experimental group participated in a concurrent task management training course and the control group did not. A Mann-Whitney U test rejected the null hypothesis, indicating a positive training effect for experimental group pilots. These results support training multi-tasking skills in single-pilot operations.

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

Effect of Concurrent Task Management Training on Single Pilot Task Prioritization Performance

Pilots must perform multiple tasks simultaneously during normal and emergency operations. They must continuously assess, prioritize, execute, monitor, and shed tasks, often in time-critical situations and in a dynamic environment. Multitasking has been recognized as a key element to successful performance in complex systems (O'Hare & Roscoe, 1990; Raby & Wickens, 1994; Wickens 1980, 1992, 2002). Three basic theories of multitasking and task management behavior that have evolved from research in cognitive psychology are 1) single channel theory (SCT), which posits that a person will abandon one task until the other task is completed (Broadbent, 1958, Welford, 1967); 2) single resource theory (SRT), or the ability to share resources between two tasks simultaneously (Kahneman, 1973; Lindsay & Norman, 1972); and 3) multiple resource theory (MRT), which assumes parallel or concurrent processing of two or more tasks is possible (North, 1977; Wickens, 1980).

The ability to effectively prioritize tasks for attention is an important flying skill and is a primary component of concurrent task management (CTM) as defined by Funk (1991) and Funk et al. (2003). According to Funk et al. (2003), CTM is the process by which pilots selectively manage concurrent tasks by assessing and prioritizing them, allocating resources in order of priority, and continuously updating their prioritization scheme to complete the flight mission safely and effectively. A task prioritization error occurs when a pilot gives preferential attention to a lower priority task rather than to a task that should take higher priority with regards to flight safety (e.g., it is more critical, more urgent, or not being performed satisfactorily) (Funk, 1991; Funk et al., 2003). Two studies conducted by Funk et al. (2003) showed that task status and procedure, such as use of checklists and cockpit flow checks, were the two main factors pilots used to prioritize tasks. Although pilots generally practice CTM, there are many instances in which failure to properly prioritize tasks or otherwise manage them effectively has led to a potentially dangerous incident or even a fatal accident (Chou, Madhaven, & Funk, 1996; Damos, 1997; Dismukes, Loukopoulos, & Jobe, 2001; Latorella, 1996; Raby & Wickens, 1994; Rogers, 1996; Schutte & Trujillo, 1996).

Task prioritization performance is inversely related to pilot workload (Chou, Madhaven, & Funk, 1996; O'Hare & Roscoe, 1990; Wickens, 2002; Wickens, Dixon, & Chang, 2003) and short-term memory appears to be a major limiting factor in CTM performance (Funk et al., 2003). Therefore, it is not surprising that computational aid to augment human memory facilitated CTM performance in two low-fidelity flight simulator experiments (Funk & Braune, 1999). However, Funk et al. (2003) listed several limitations to developing and employing CTM aiding systems in the cockpit, including unreliability of speech recognition technology in a demanding cockpit acoustic environment, the complexity of the software, certification challenges, and the interface of the systems with existing aircraft avionics. Funk et al. (2003) suggested that other means of improving CTM be explored, notably the training of CTM skills, including that of prioritizing tasks.

Bishara and Funk (2002) developed and evaluated a short (two-hour) CTM training module for general aviation pilots. No significant differences in task prioritization performance were found between the group of participants who received CTM training and the control group. This may have been due to several factors, including the quality of the training material (not developed by qualified flight instructors), the low fidelity of the simulator (a keyboard and mouse were used for virtual aircraft control), a small sample size (12), and the heterogeneity of the participants, who varied widely in experience, pilot certification, and flight time (Funk et. al., 2003). Although CTM performance is a significant factor in flight safety, the trainability of CTM, until now, has been in question.

Objectives

The main objective of this study was to design and implement a CTM training course using established educational practices and FAA and industry training methods and to determine its effect on single pilot task prioritization performance in simulated flight. Additionally, this experiment was conducted using a higher fidelity flight simulator platform and a larger and more homogeneous group of pilots than previous studies. A secondary objective was to observe pilot behaviors related to multitasking theories, including cognitive processing models of SCT, SRT, and MRT.

Method

A pretest-posttest control group design with random assignment was used. All participants flew a one hour simulated instrument flight on a Frasca 141 flight training device FTD (pretest) then flew another simulated flight two weeks later (posttest). The experimental group participated in a CTM training course during the two week interim period and the control group did not.

Task prioritization challenges and associated errors were defined based on CTM theory developed by Funk et al. (2003) and on data gathered during a pilot study prior to the experiment. Pilot study data were used to establish procedures and define the method used to evaluate task prioritization errors. Each pilot's error scores were coded by live observer as well as video taped and coded by a second observer using a blind procedure; none of the coders knew which pilots were in which group. To minimize contamination all participants, instructors who conducted the FTD flights, and coders were instructed to avoid discussing the experiments with anyone else until after the study was completed.

Participants

Twenty-seven pilots enrolled in the Central Washington University (CWU) Flight Technology Program participated in the experiment. The majority of participants were enrolled in the intermediate instrument stage of training, with the exception of two who were certified flight instructors. All pilots had logged previous instrument time on the Frasca 141 FTD used in the experiment. Two of the participants were women, and the remaining 25 were men. Participants ranged in age from 19 to 22. They were randomly assigned to the control group (14 participants) and the experimental group (13 participants). Pilots reported their flight experience, including instrument and FTD time.

Flight Training Device

Two identical Frasca 141 FTDs were used and were configured as normally aspirated single-engine fixed-gear aircraft using the same performance parameters with which participants were accustomed in flying airplanes and FTDs during prior training. The avionics package included a Bendix/King stack with dual KY196 Communication radios, dual KN53 navigation radios, KDI 572 DME, KR 87 ADF, KT76A transponder, KMA 24 audio panel with marker beacons, and GNS430 IFR enroute and approach certified GPS. The FTDs recorded all primary flight data including aircraft heading, altitude, airspeed, power settings, and position.

Procedure

Pilots conducted simulated flights as per the CWU Standard Operating Procedures (SOP) manual; all checklists, flow checks, and callouts were the same used in their normal flight training. Pre- and posttest simulated flights were conducted in a line oriented flight training (LOFT) format. The same LOFT was flown for both flights; it placed pilots in a high workload environment in Seattle Class B airspace and included radar vectors as well as pilot navigation, two precision instrument approaches, a multistage missed approach, and a holding procedure.

Certified instrument flight instructors (CFIIs) were trained to administer the LOFT which was scripted with respect to air traffic control (ATC) communications and procedures. Flights were observed and scored in real time and again from videotape by a second scorer. Video cameras recorded a wide angle view that included the entire instrument panel, engine controls, yoke, rudder pedals, and pilots' hands and feet.

Practice effects were minimized partly by the condition that all pilots had previous time on the Frasca 141 FTD and should not have been learning how to operate the FTD during observation flights. Comparison of pretest and posttest error data from the control group did not indicate a practice effect due to repeating the same LOFT scenario for both flight trials. In fact, several pilots in the control group made more errors in the posttest flight than they did in the pretest. Also, the specific errors and types of errors they made in the posttest were not the same as those in the preflight. For example, a pilot from the control group who made the same number of errors in the posttest as in the pretest did not in general make the same types of errors or the same errors, which indicated there was probably not a practice effect due to pretest sensitization.

Prioritization Scheme

A task prioritization scheme taught to pilots during primary and advanced flight training is the aviate, navigate, communicate (ANC) hierarchy (Chappell, 1998; FAA, 1999; Jeppesen, 2002, 2003a, 2003b; Kern, 1998; Kershner, 1998; Machado, 2001, 2003; Thom, 1991). For this study, each task was defined based on current pilot training manuals and literature as follows:

- *Aviate task:* Included all items related to aircraft operation: airspeed, altitude, climb or descent rate, and changes in lift, thrust, and drag; e.g. primary aircraft control inputs (pitch, power, yaw, and roll), operation of lift and drag devices (flaps) and operation of primary engine systems.
- *Navigate task:* Included items related to the current and future position of

the aircraft, including vectors, course intercepting and tracking, identification of intersections and waypoints, and programming and operating the GPS and other navigation radios.
- *Communicate task:* Included communications with ATC.

Definition of Task Prioritization Errors

Opportunities for 20 potential task prioritization errors were embedded at 14 challenge points throughout the one-hour simulated flights. Similar to the method used by Funk et al. (2003), each challenge point provided an opportunity for the participant to divert his/her attention from a more important or more urgent task to a less urgent or less important task. Specific challenges were based on errors observed during a pilot study conducted prior to the experiments and each challenge point specified what actions would constitute which type of prioritization error. Types of prioritization errors included ignoring an aviate (flight control) task in order to navigate (aviate/navigate, 7 opportunities), aviate/communicate (7), navigate/communicate (5), and aviate/aviate (1) in which the pilot had to choose between two aviate tasks as to which was most critical to perform first.

Several of the challenge points were simply part of the LOFT scenario; they were embedded at a point where a pilot might make a task prioritization error and thus did not require any intervention. For example, challenge points were placed at locations in the flight scenario where there was potential for error if the pilot fixated on or became distracted by a navigate task at the expense of primary aviate tasks. Other challenge points required the CFII to act as ATC and call the pilot with information or instructions just before the pilot was leveling off or about to intercept course, or to cause a failure to a navigational facility or an aircraft system.

Performance criteria for determining if an error occurred was based on FAA-S-8081-4C Instrument Rating Practical Test Standards with respect to altitude, airspeed, heading, intercepting and tracking course, use of checklists, procedures, and ATC communications.

CTM Training Course

The training course followed standard practices and procedures common to the CWU FAA approved training course outline (TCO) and university-approved criteria for learning outcomes and assessment strategies. The course was taught by an FAA-certified CFII and CWU flight technology professor and consisted of two learning sessions 10 days apart that included reading, self-study, cooperative learning activities, guided discussion, and a reflective homework assignment. Funk et al. (2003) determined that procedure was the primary reason pilots prioritized tasks, which corroborated analyses by Chappell (1998) and Kern (1998). Therefore, the training course emphasized procedural discipline, including adherence to standard operating procedures, conduct of checklists, briefings, flow checks, and mnemonic memory aids at appropriate times. The training course did not stress any one type of task prioritization error over another but rather addressed task prioritization concepts with respect to various scenarios and possible types of errors related to improper use of checklists or lack of procedural discipline, as well as focusing on the ANC scheme to assess what took priority in a given situation.

The first learning session consisted of a class discussion of selected materials related to aviation human factors, aeronautical decision making, situational awareness, workload management, and concurrent task management. Participants had prior knowledge of all those concepts from previous coursework and studies, thus the training did not introduce any new concepts but rather emphasized task prioritization as an important element of human factors and aeronautical decision making. Participants analyzed accident and incidents taken from the National Transportation Safety Board (NTSB) and National Aeronautics and Space Administration (NASA) databases with respect to CTM errors, including task prioritization errors, and participated in class discussions of those data.

Between sessions, participants were asked to reflect on at least one of their normal flights with respect to CTM concepts and how their awareness influenced their in-flight decision making. Students reflected in writing as well as through a verbal debriefing.

The second learning session included an activity in which participants acted out role-playing scenarios designed to give insight into their reactions and behavior in the cockpit when confronted with CTM challenges. They also participated in a class discussion of strategies to improve pilot task prioritization performance and a guided discussion of the outcomes. A short quiz was given at the end of the second session to evaluate each pilot's progress and identify areas of improvement. The CTM training course emphasized using low workload times to conduct non-critical tasks ahead of time and reminded pilots to continuously assess the status of the aircraft and prepare for the next phase of the flight. The concept was mainly stressed through procedural discipline and strict use of checklists and flow checks.

Results

Using the bivariate coefficient of determination interpretation of Newton and Rudestam (1999), there was little to no correlation between participants' total flight time, instrument time, stage of training, and Frasca 141 FTD time with regards to task prioritization performance on the pretest, and correlation between total FTD time and pretest prioritization error scores were negligible to weak (Table 1). An independent samples t-test comparing pretest mean error scores for the two groups indicated there was no significant difference between the two groups' pretest error scores ($t = 0.14$, $df = 25$, $p = 0.88$). These results revealed that the two groups were equivalent with respect to their pretest task prioritization performance. Thus, any large variations in posttest prioritization error scores could more likely be attributed to the independent variable (training course) rather than to pretest differences between groups.

Table 1
Coefficient of determination for pilot criteria and pretest task prioritization error scores according to group. The coefficient of determinations showed correlations were negligible to weak, indicating the two groups could be considered equivalent and relatively homogeneous with respect to pretest variables.

Coefficient of determination (R^2) for pilot criteria and pretest CTM error scores					
	Total time	Stage of training	Instrument time	FTD time	Frasca 141 time
Control group Pretest scores	0.036	0.018	0.002	0.001	0.097
Experimental Group pretest scores	0.148	0.021	0.043	0.202	0.106

Task prioritization error data were recorded as a frequency distribution of raw scores and converted to a ratio score (number of errors:total number possible). Comparison showed large pretest-posttest differences for the experimental group and little change for the control group (Figure 1). The experimental group showed a 54% decrease in total errors from pretest to posttest and the control group showed a 9% increase in total task prioritization errors.

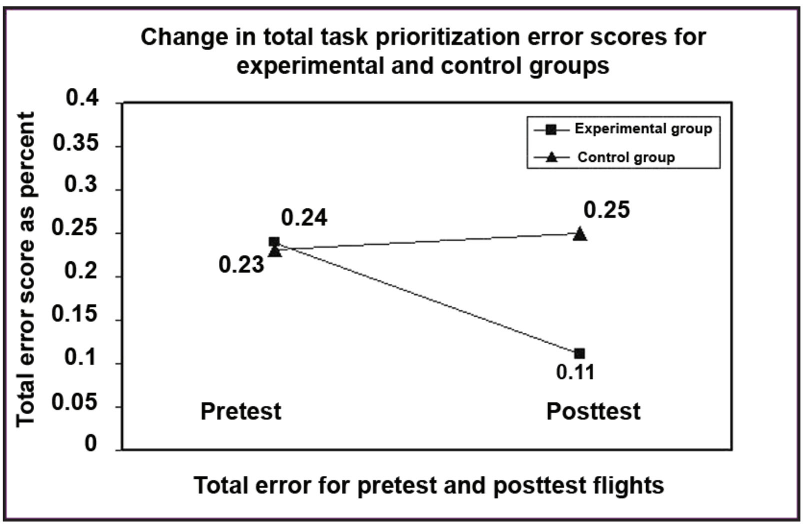


Figure 1. Change in total task prioritization error scores for experimental and control groups expressed as a percent of total possible errors.

An F-test for homoscedasticity found the samples had equivalent variance and a K-S test and Q-Q plot showed they were normally distributed. However, a histogram of the distribution showed a slight increase in error scores at both tails. Because parametric statistics that compare means (such as a t-test) are more sensitive to deviations at the tails of a distribution, a more conservative nonparametric statistic was chosen and a Mann-Whitney U test was used to compare the task prioritization performance for the two groups (Table 2). The Mann-Whitney U test was calculated as described by McClave and Sincich (2003) and rejected the null hypothesis that there was no difference in posttest errors between the groups (p = 0.029), thus indicating a positive training effect for experimental group pilots.

Table 2
Results from Mann Whitney U test

Mann-Whitney U test - Not corrected for ties	
	pretest-posttest difference
U	46.0
U critical	50
Z	-2.2
p (2-tailed)	0.029

The distributions of types of pretest and posttest errors are shown in Figures 2 through 5. Because there was only one A/A error challenge point in the LOFT, and it represented only 1.0% of total errors in the pretest and 0.7% of total errors in the posttest, it was not included in the graphs of error score breakdowns into type. Comparison of pretest error scores (Figure 2) shows the experimental group had 5% fewer A/N errors, 10% more A/C errors, and 30% more N/C errors than the control group. Comparison of posttest error scores (Figure 3) shows the experimental group had 47% fewer A/N errors, 56% fewer A/C errors, and 53% fewer N/C errors than the control group. For both groups an A/N error involving fixation on the GPS as discussed in the following section was the most common, followed by an A/C error involving an interrupting ATC call just as pilots were leveling off (A/C challenge) and as pilot intercepted a localizer course (N/C challenge).

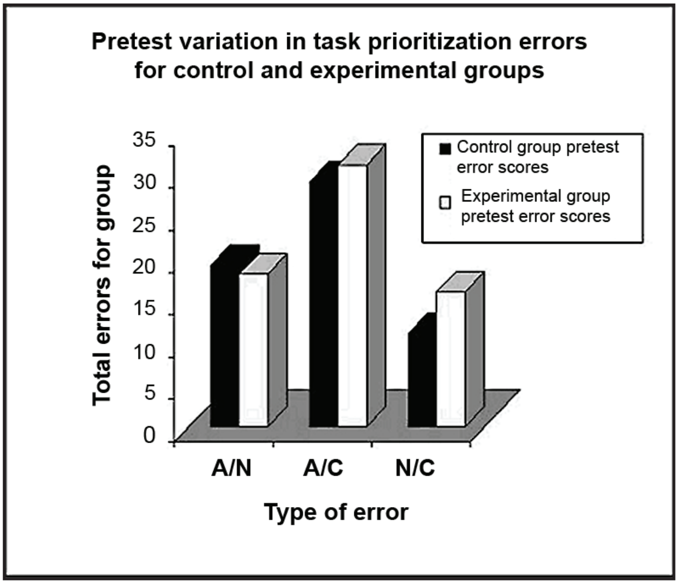


Figure 2. Pretest variation in task prioritization errors for control and experimental groups. The experimental group had 5% fewer A/N errors, 10% more A/C errors, and 30% more N/C errors than the control group.

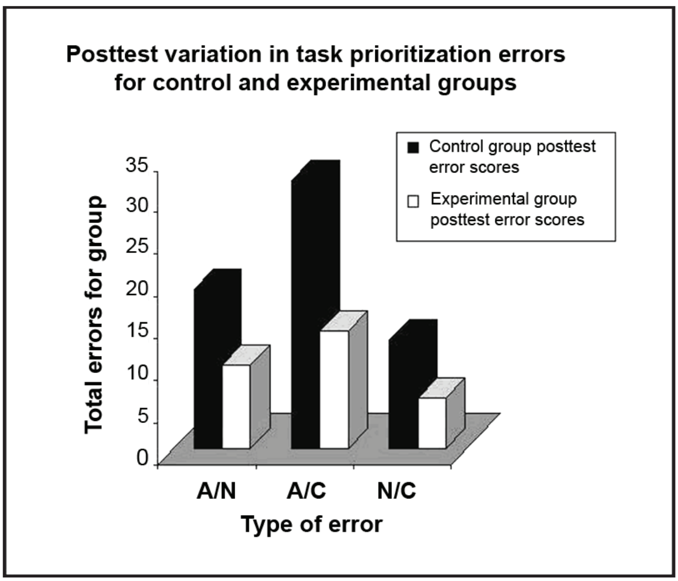


Figure 3. Posttest variation in task prioritization errors for control and experimental group. The experimental group had 47% fewer A/N errors, 56% fewer A/C errors, and 53% fewer N/C errors than the control group.

The distribution of errors for the control group shows there was not a significant change in any one type of error from pretest to posttest (Figure 4). The group made the same number of A/N errors in both the pretest and posttest flights, A/C errors increased 9% and N/C errors increased 15%. With only one or two exceptions there was no relationship between the types of errors made by a given control group pilot in both pretest and posttest. As a whole the experimental group showed a 44% decrease in A/N errors, a 55% decrease in A/C errors, and a 63% decrease in N/C errors (Figure 5).

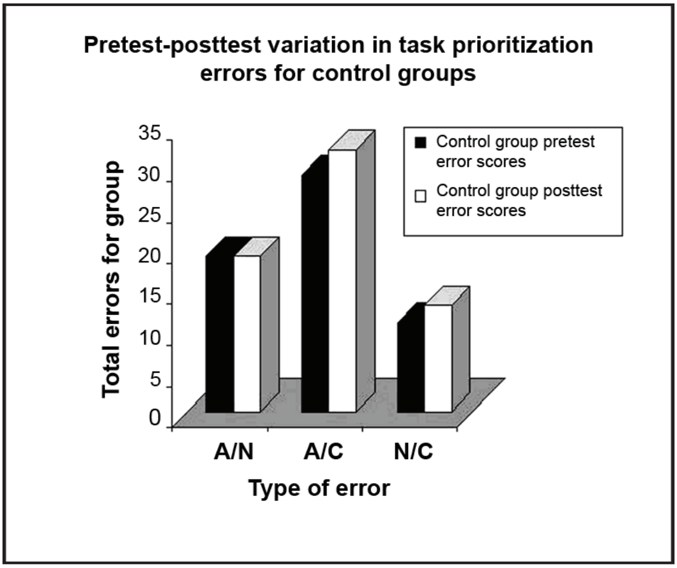


Figure 4. Pretest-posttest variation of each type of task prioritization error for the control group.

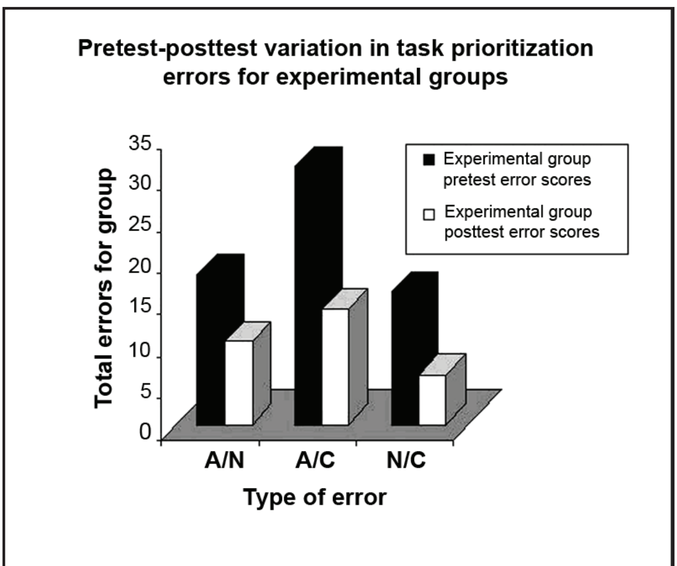


Figure 5. Pretest-posttest of each type of task prioritization error for the experimental group.

Discussion

A positive short term training effect for experimental group pilots was observed. Several pilot multitasking behaviors could be described by SCT, SRT, and MRT indicating the efficacy of CTM training, as described below. Concepts of CTM theory related to procedural discipline and workload management are also discussed.

Training Effect

The experimental group showed a large decrease in total task prioritization errors between pretest and posttest flights compared to the control group (Figure 1). Pilots in the control group made the same or more prioritization errors overall in the posttest flight compared to the pretest, which could be expected from randomly sampling a group of pilots during two discrete flights. If there were no effect from the CTM training course then the experimental group should show a similar distribution of pretest and posttest scores. It follows that the CTM training course may have caused a reduction in task prioritization errors for pilots in the experimental group over the two week period of time evaluated by this experiment. However, the training effect on longer term learning was not evaluated.

One of the most widely used theories of learning and retention is the dual memory model described by Schunck (2003) and adopted by the FAA in their flight instructor training literature (FAA, 1999). According to the dual memory model, information is processed through inputs (primarily visual and auditory) to the sensory register. In order to transfer information to long term memory the learner must relate incoming information to concepts and ideas already in memory. Pilots in the experiment had previously studied concepts of prioritization and task management during their regular flight training, and it is possible that the reduction in task prioritization errors by the experimental group might represent a sensitization effect; the only difference between the two groups might have been that the experimental group was focused on those concepts during the short term and did not actually code the information into their long term memory. The issue of whether long term learning occurred is a critical one and also difficult to resolve because a teacher or instructor often does not have the ability to evaluate students after they leave the learning environment. More follow-up studies are needed to evaluate long term effects of the training. Additionally, a qualitative response from participants at some future time might also reflect on whether or not they believed learning occurred.

Pilots in the experimental group who showed the greatest reduction in error scores were the ones that originally made the most errors. Thus it could be that the reduction in errors might simply represent a regression toward the mean for those pilots. However, the fact that several pilots in the control group also scored a large number of errors in the pretest without a corresponding reduction in errors for the posttest indicated that regression was probably not the cause for that trend in the experimental group. What the data suggested was that pilots who performed the worst seemed to benefit more from the training than those who initially made a low number of errors. Alternatively, this may represent a ceiling effect for pilots who made only one or two errors in the pretest, since there were only a fixed number of challenge points.

As part of the CTM training course, pilots in the experimental group wrote a reflection of their perceived task prioritization performance during their most recent flight and some debriefed that flight verbally with the investigator. Many pilots noted that they had a heightened awareness of the highest priority item, flying the airplane, during those flights; but they did not change the way they conducted flow checks or checklists. Rather, the training reaffirmed their knowledge and heightened their awareness of aviate tasks as being the most important. From an educational standpoint, the hope is that because pilots in the experimental group were given the opportunity to reflect and apply concepts from the CTM training course they were able to process and retain the information and that learning occurred. However, further studies are needed to confirm longer term training effects.

Multitasking Behaviors

The challenge points in the LOFT were designed to simulate situations that might occur in a routine flight where pilots were confronted with two or more simultaneous tasks that had to be prioritized. Following is a discussion of how pilot behaviors in this experiment might be explained by three different SCT, SRT, and MRT cognitive processing models.

Single channel theory. To resolve some of the LOFT challenges, some pilots made prioritization errors that could be described by SCT. For example, more than half the pilots in both groups made an A/N error that involved a missed approach procedure (MAP) which required the pilot to climb via the localizer course to 2000 feet, then to identify a specific intersection as the point to commence climb to 5000 feet while continuing to track the localizer course. Many of the pilots became fixated on the task of either programming the GPS for the waypoint or tuning and identifying the VOR to identify the cross radial for the intersection and either strayed off course, deviated from altitude, or both, while attempting to identify the fix. In several cases the video tapes showed pilots were not even looking at their flight instruments but had leaned over to the right side of the cockpit to view and input the GPS unit more directly. A few pilots were off altitude by as much as 500 feet, off the localizer course by a full needle deflection, or off heading by as much as 70 degrees within a matter of only a few seconds. Additionally, some of those pilots did not immediately recognize their altitude and course/heading deviations when they looked back at the flight instruments. Moray and Rotenberg (1986) called that phenomenon "cognitive lockup," which in the situation just described meant the pilots became attentionally locked into the navigate task and did not notice the large deviations in primary aircraft control or navigation tasks. Moray and Rotenberg concluded that cognitive lockup behavior represents evidence that people deal with problems serially rather than switching between tasks. The type of fixation exhibited by those pilots can be described by SCT. Although pilots in the study were familiar with the GPS unit from previous use in both the airplane and FTD, they still became fixated when programming the waypoint in flight. Pilots who pre-programmed the GPS while still on the ground did not make that fixation error.

The issue of fixation has become an area of great concern in the flight training industry in recent years. Over the past five years, general aviation cockpits have incorporated more sophisticated IFR certified GPS units and flat panel primary and multifunction displays, and currently new training aircraft are delivered with all glass cockpits. Wilson and Funk (1998) found that as the level of sophisticated instruments and automation increases on airline flight decks the potential for task prioritization errors also increases. Also, in a more general meta-study of airline flight deck human factors issues, Funk and Braune (1999) found the attentional demands of automation to be problematic. It is likely that the same potential exists for increased sophistication in general aviation cockpits, including training aircraft.

Single resource theory. Observations indicated some of the pilots showed an ability to share resources between two tasks simultaneously as described by SRT. Wickens (2002) posited that mental resources can be allocated to different tasks and if one is simple (or automatic) it requires almost no mental resources, so more of the available resources can be allocated to the other task. An example of that behavior was observed at a specific challenge point which required the pilot to execute a MAP. During a MAP the most critically important aviate task is to initiate a positive rate of climb before doing any other task. Participants in this study had been previously trained to use a mnemonic memory aid for the MAP called "five Cs," which is designed to guide them in doing tasks in the order of higher to lower priority. The five Cs stands for "cram, climb, clean, cool, call" and means the pilot should cram the throttle (put the throttle, and propeller if applicable, to the full forward position), initiate a positive rate of climb, clean up the airplane aerodynamically (retract the landing gear and flaps), cool the engine via opening of cowl flaps or richening of fuel mixture, and call ATC as last priority.

The first two actions (pushing the throttle forward while pitching for best rate of climb attitude) were done simultaneously by most of the pilots. According to SRT, the time for the pilot to execute both those tasks simultaneously will be less than the time required to do them sequentially because the pilot is able to perform the more automatic task (increasing the throttle) without using much mental effort so they can concentrate more on the more difficult task of pitching to the correct attitude on the attitude indicator. According to Wickens (2002) there is some task interference, but it results not from postponement of one task over another (as predicted by SCT) but rather from the concurrence of the tasks. Wickens stated that both tasks will probably be executed in the amount of time of the longest task; e.g. if it takes three seconds to push the throttle forward and seven seconds to pitch for climb attitude, the total time for both tasks will be seven seconds (the time it takes for the more difficult task). SCT would predict a total time of 10 seconds for the pilot to execute the tasks sequentially. Because most pilots did both tasks simultaneously in a relatively short time they were most likely sharing resources between the tasks according to SRT.

An exception was one pilot in the control group who did not use the memory aid and made the same prioritization error during both the pretest and posttest flights. Instead of putting the throttle to full power and initiating a climb, the pilot put the throttle to about half power, stopped and called ATC, then put the throttle to about the three-quarter position. Then he stopped and switched radio frequencies (during that time in the pretest flight the climb degraded and the plane went into a descent, in the posttest it leveled off), after which he called ATC, then finally put the throttle to full forward and established a climb. The pilot was having great difficulty switching between tasks and did not seem to be able to timeshare between tasks. In fact the whole sequence took over two minutes in the pretest and almost the same amount of time in the posttest before he actually established a positive rate of climb. Other pilots typically achieved a positive rate of climb within 10-15 seconds.

The MAP scenario just described is probably one of the most critical to flight safety, since on a typical MAP from an ILS the aircraft is only 200 feet above the ground, and if a climb is not initiated immediately there is a high risk for a controlled flight into terrain (CFIT) accident. Because the pilot just discussed made the same error in a similar manner both times, it could represent a learned behavior that would require significant additional training effort to rectify, or it could represent an inability to cognitively process the tasks, which could be potentially dangerous.

Multiple resource theory. Some pilots exhibited an apparent ability to process two or more tasks concurrently as described by MRT. According to Liu and Wickens (1992) and Wickens, Dixon, and Chang (2003), tasks that do not compete for the same resources, such as a visual task and an auditory task, are easier to perform simultaneously than two tasks that use the same resources (e.g.: two visual/spatial tasks). Observations from this study confirmed that the visual and auditory tasks interfered less with one another than did two visual tasks. At one particular challenge point, pilots had to prioritize between a higher priority aviate task (visual) and a communicate task. Pilots had to visually interpret inputs from the primary flight instruments to level the airplane at a designated altitude while simultaneously attending to an incoming radio call from ATC (auditory task). Another challenge point was similar, except that the visual input was a navigate task (interpreting movement on the VHF navigation display and intercepting the localizer course) and the communicate task again involved attending to an incoming radio call from ATC (auditory task). Most pilots were able to level off the airplane or make the turn to intercept and attend to the incoming call at the same time, thus effectively time sharing their attention between tasks that required different cognitive resources (visual and auditory) as suggested by MRT. A few pilots postponed the communicate task until after completing the aviate or navigate task, which could indicate they were processing inputs in a more sequential manner that would be consistent with SCT.

In contrast, another challenge point involved a conflict between a visual aviate task (flying the aircraft at a constant airspeed and descent rate) and a visual navigate task (tracking the localizer and determining that the glideslope had failed). According to MRT, two tasks that use the same resources are more difficult to perform simultaneously. Indeed, twice as many pilots made the error that involved two visual tasks compared to those that made errors involving a visual and an auditory task, as just described. MRT could also explain the situation during the MAP described in the section on SCT. The videotapes clearly showed that many pilots completely ignored the aviate task to lean over and focus on the GPS, as described earlier. However, some of the pilots could have been unsuccessfully attempting to time share or switch between tasks, which would be better described by MRT rather than SCT.

Concurrent Task Management

Because a substantial amount of research exists regarding task management and concurrent task management (Funk, 1991; Funk et al., 2003; Raby & Wickens, 1994; Rogers, 1996; Schutte & Trujillo, 1996; Wickens et al., 2003) this study used criteria from previous studies to design elements of the CTM training course and to note how particular task prioritization error performance changed between pretest and posttest flights. Funk et al. (2003) showed that procedure and tasks status were the two primary factors pilots identified as influencing the way they prioritize cockpit tasks, and the CTM training course focused on procedure as described earlier.

The LOFT placed the pilot in a relatively high workload environment at different times throughout the flight. Pilots who completed as many tasks as possible ahead of time during times of relatively low workload had better task prioritization performance during the high workload times (departure, approach, and missed approach). For example, pilots who pre-programmed the GPS on the ground before departure did not make the fixation error during the MAP as previously described. Likewise, pilots who obtained weather updates well prior to commencing the approach, and who briefed the approach or hold well ahead, had more time for the actual conduct of those maneuvers in addition to being able to monitor flight and engine instruments. Those observations corroborated empirical results of Raby and Wickens (1994), which showed that pilots who conducted tasks during low workload periods performed better.

Conclusions

Experimental analysis showed that the group of pilots who participated in the CTM training course improved their task prioritization performance over a two-week period of time. Those pilots showed improvement in each of the three main types of prioritization errors. It was not determined whether that performance increase had a longer lasting effect. Pilots who did not participate in the CTM training course showed either an increase, decrease, or no change in their task prioritization performance. Based on the control group's posttest performance, there did not seem to be a practice effect from the pretest.

Pilots in the experimental group who made the most task prioritization errors in the pretest showed the most improvement after participating in the CTM training course. However, that decrease in errors did not seem to be a result of regression toward the mean since control group pilots with comparable pretest performance showed negligible change in posttest flights. Experimental group pilots who made only one or two errors in the pretest might have improved or not, but that criteria was undeterminable since there were a fixed number of challenge points from which data were collected instead of all assessing all possible errors made.

Pilot behavior was described by different cognitive processing models with regards to various behaviors and task prioritization errors. Some behaviors could be described by SCT, wherein the pilot did not seem to be able to process inputs and perform tasks simultaneously. SRT described other behaviors in which pilots seemed to be able to time share or switch between tasks. As described by multiple resource theory MRT, tasks that shared the same cognitive resources, such as two visual tasks, were not performed as well together as tasks that did not, such as a visual task and an auditory one. One particular error that emerged was that of pilots fixating on the GPS display to the exclusion of aircraft control, sometimes showing dangerously large deviations in altitude and course. Fixation errors are of critical importance in the current flight training environment as modern cockpits utilize more sophisticated displays and avionics.

Pilots who practiced a high level of procedural discipline and proper use of checklists, flow checks, and mnemonic memory devices performed better than those who did not, especially when they were able to perform routine tasks during low workload periods. Experimental group pilots who improved their procedural discipline and performed more tasks during low workload periods showed a marked improvement in task prioritization performance in posttest flights.

Limitations

Although recommendations from previous studies were addressed by this experimental analysis, other items were considered to be limiting factors. The time period between pre and posttest FTD flights was a compromise between internal and external validity; the two-week duration was short enough to increase internal validity because it minimized possible history effects from extraneous learning, but it could also have made the results less meaningful with regards to the long term effects of training. Control for extraneous variables over a longer period of time, including further training in human factors and additional flight experience, might be challenging. However, pretest data indicated no relationship between this particular group of participants' total flight time, instrument time, or FTD time and their task prioritization performance, so controlling for the influence of such extraneous variables could be a reasonable possibility for similar groups. Future studies are needed to examine longer term training effects and to investigate how long it takes for any training effect to disappear or to drop below acceptable performance standards.

Selection of the participants was another compromise between internal and external validity. Participants comprised a relatively homogeneous group with respect to previous training and experience, but those same criteria means that results may not be generalized to a more variable group of general aviation pilots. Thus, similar studies with other pilot groups are needed.

FTDs were fully FAA approved and equipped with standard analog flight and engine instruments as well as standard avionics. The Garmin GNS430 approach certified GPS was a more advanced system. Because fixation on the GPS was a major cause of task prioritization error for more than half the pilots, any experiment such as this one using basic analog instrumentation might not address the challenges faced by pilots using the newest flat panel digital PFD and MFD displays now being installed in many new aircraft. This too represents another area for future research.

The training course was carefully designed and delivered to be consistent with current practices used by FAA certified flight instructors and flight schools. However, the instructor who conducted the course has a substantial amount of experience teaching in the classroom and in airplanes and simulators, as well as experience training flight instructors. Therefore, the experiment could have been only testing one specific instructor and thus not be able to generalize results to other instructors or learning environments.

Recommendations

Based on the limitations just discussed, several recommendations for future research are suggested. First, the same experiments could be conducted with a different group of pilots. External validity issues could be addressed by comparing results between pilots who have different training backgrounds. If results were similar, then the training course could be more generalized to other training programs. Additionally, the training course should be conducted by other flight instructors with varying levels of experience so that the results can be correlated as to whether or not they were dependent on a particular instructor conducting the training. A less homogeneous group of pilots might also be used to investigate whether pilots with varying levels of experience showed different training effects and if there were some specific level of experience at which pilots showed the greatest effect. External validity could also be enhanced by using a larger sample size and a more powerful experimental design, such as a Solomon four-group. Other experimental designs, such as a time-series design, could be used to determine longer term training effects, either with single subjects or a group of pilots.

Experiments could be conducted with a longer time period between pretest and posttest flights and controlled for extraneous variables to test for long term training effects. Qualitative studies could also be used to enhance experiments, such as gathering responses from participants to discern the extent of their learning.

To investigate the issues of task fixation as it relates to cockpit complexity and automation, a study could be designed to test pilots in cockpits with various levels of complexity; for example, using one of the many new flat panel PFD/MFD or virtual 3D displays installed in many new general aviation aircraft.

Acknowledgements

This study was funded in part by a grant from the Constance Wolf Aviation Fund. The authors would like to thank students and instructors at Central Washington University for their participation.

References

Bishara, S. & Funk, K. (2002, September 30 – October 4). Training pilots to prioritize tasks, *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, Baltimore Maryland, 96-100.

Broadbent, D. (1958). *Perception and communications.* New York: Permagon Press.

Chappell, S. L. (1998). Disciplined attention, the next best thing to a crystal ball. In T. Kern (Ed.). *Flight discipline*. New York: McGraw-Hill.

Chou, C.D., Madhavan, D., & Funk, K.H. (1996). Studies of cockpit task management errors, *International Journal of Aviation Psychology*, *6*(4), 307-320.

Damos, D. L. (1997*).* Using interruptions to identify task prioritization in part 121 air carrier operations. *Proceedings of the Ninth International Symposium on Aviation Psychology(4),* Columbus, Ohio State University.

Dismukes, R. K., Loukopoulos, L. D., & Jobe, K. K. (2001). The challenges of managing concurrent and deferred tasks. In R. Jensen (Ed.), *Proceedings of the 11th International Symposium on Aviation Psychology.* Columbus: Ohio State University.

Federal Aviation Administration (1999). *Aviation instructor's handbook* (FAA-H-8083-9). Washington D.C.: U.S. Government Printing Office.

Funk, K. H. (1991). Cockpit task management: Preliminary definitions, normative theory, error taxonomy, and design recommendations. *The International Journal of Aviation Psychology. 1*(4): 271-285.

Funk, K. & Braune, R. (1999, October 19-21). *The AgendaManager: A knowledge-based system to facilitate the management of flight deck activities*, SAE 1999-01-5536, SAE International, Warrendale, PA, presented at 1999 World Aviation Congress, San Francisco, CA.

Funk, K. H., Colvin, K, Bishara, S., Nicolalde, J., Shakeri, S., Chen, J. Y., & Braune, R. (2003). *Training pilots to prioritize tasks: Theoretical foundations and preliminary experiments.* NASA Grant NAG 2-1287 Final Report. Corvallis, Oregon.

Jeppesen (2002). *Flight instructor manual*. Englewood, CO: Author.

Jeppesen (2003a). *Instrument commercial manual*. Englewood, CO: Author.

Jeppesen (2003b). *Private pilot manual.* Englewood, CO: Author.

Kahneman, D. (1973). *Attention and effort*, Englewood Cliffs, NJ: Prentice-Hall.

Kern, T. (1998). *Flight discipline*. New York: McGraw-Hill.

Kershner, W. K. (1998). *The instrument flight manual* (5th ed.). Ames, IA: Iowa State University Press.

Latorella, K. A. (1996). Investigating interruptions: An example from the flight-deck. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting.* 249-253.

Lindsay, P. H., & Norman, D. A. (1972). *Human information processing*. London: Academic Press.

Liu, Y. & Wickens, C. D. (1992). Visual scanning with or without spatial uncertainty and divided and selective attention. *Acta Psychologica, 79*, 131-183.

Machado, R. (2001). *Rod Machado's private pilot handbook.* San Clemente, CA: The Aviation Speaker's Bureau.

Machado, R. (2003). *Rod Machado's instrument pilot's survival manual*. San Clemente, CA: The Aviation Speaker's Bureau.

McClave, J. T., & Sinich, T. (2003) *Statistics.* (9th ed.). Upper Saddle River, NJ: Prentice Hall.

Moray, N. & Rotenberg, I. (1986) Fault diagnosis in process control. *Annual Proceedings of the Human Factors Association of Canada,* Montreal.

Newton, R. R., & Rudestam, K. E. (1999). *Your statistical consultant: Answers to your data analysis questions.* Thousand Oaks, CA: Sage.

North, R. A. (1977), Task functional demands as factors in dual-task performance. *Proceedings of the 21st Annual Meeting of the Human Factors Society.* Santa Monica, CA.

O'Hare, D., & Roscoe, S. (1990). *Flightdeck performance: The human factor.* Ames, IA: Iowa State University Press.

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *The International Journal of Aviation Psychology. 4*(3): 211-240.

Rogers, W. H. (1996). Flight deck task management: A cognitive engineering analysis. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting,* 239-243.

Schunk, D. H. (2003). *Learning theories: An educational perspective* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Schutte, P. C., & Trujillo, A C. (1996). Flight crew task management in non-normal situations. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting.* 244-238.

Thom, T. (1991). *The pilot's manual: Volume 2 flight operations.* New York: Center for Aviation Theory.

Welford, A. T. (1967) Single channel operation in the brain. *Acta Psychologica, 27*, 5-21.

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), *Attention and performance VIII*, (pp. 239-257). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper Collins Publishers.

Wickens, C. D. (2002), Situation awareness and workload in aviation. *Current Directions in Psychological Science, 11*(4)*,* 128-133.

Wickens, C. D., Dixon, S., & Chang, D. (2003). *Using interference models to predict performance in a multiple-task UAV environment – 2 UAVs* (Tech. Rep. AHFD-03-9/MAAD-03-1). Urbana-Champaign, IL: University of Illinois Aviation Human Factors Division Institute of Aviation.

Wilson, J. R., & Funk, K. (1998). The effect of automation on the frequency of task prioritization errors on commercial aircraft flight decks: An ASRS incident report study. *Proceedings of the Second Workshop on Human Error, Safety, and Systems Development, 6-16.*

International Journal of Applied Aviation Studies, 5, Number 2

Employee Attitudes Within the Federal Aviation Administration

Carla A. Hackworth, Crystal E. Cruz, Scott Goldman
Civil Aerospace Medical Institute
Federal Aviation Administration

Dan G. Jack, S. Janine King
Xyant Technology, Inc.

and

Paul Twohig
Office of Human Resources
Federal Aviation Administration

Direct Correspondence to:
FAA CAMI AAM-510
PO Box 25082
Oklahoma City, OK 73125
405-954-6299, 405-954-4852 (fax)
Carla.Hackworth@faa.gov

Abstract

Prompted by Congressional direction, the Federal Aviation Administration (FAA) instituted a means of assessing employee attitudes following the 1981 air traffic controllers' strike. The Employee Attitude Survey (EAS) has been administered nine times since its inception. The 2003 EAS contained 129 items organized into three major sections: (1) Indicators of Satisfaction, (2) Management and Work Environment, and (3) Respondent Demographics. The 22,720 valid surveys returned represented a 46% response rate. Seventy-one percent of respondents indicated that they were somewhat or very satisfied with their jobs. Most employees (81%) were also committed to the FAA, satisfied with their compensation (65%), and satisfied with their immediate supervisors (61%). Overall, the FAA has a committed workforce with a high level of job satisfaction. However, only 21% of FAA employees agreed that corrective actions are taken to deal with poor performers. Performance management is a common problem for many organizations. A corporate plan addressing several areas identified as needing improvement on the 2003 EAS is discussed.

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

The Federal Aviation Administration (FAA) is responsible for maintaining safe and expeditious air travel within the National Airspace System (NAS). The FAA fulfills this mission through the efforts of nearly 50,000 employees in 11 Lines of Business (LOBs)/Major Organizations (MOs; Table 1). The FAA is the largest agency within the Department of Transportation (DOT, 2003); its functions include certifying aircraft, providing air traffic control services and other assistance to commercial and general aviation pilots, and maintaining the infrastructure of the NAS (e.g., radar, towers). Managing a workforce of this size requires the efforts of many dedicated professionals, as well as multiple data sources to gauge progress and performance goal targets. One such data source is the direct feedback of FAA employees through a survey of employee attitudes.

Table 1
*FAA Lines of Business/Major Organizations**

Organization	Acronym	Description
Office of the Administrator	AOA	Executive offices, financial services, civil rights, noise and emissions control, international aviation, legal offices, public affairs, and human resource
Commercial Space Transportation	AST	Regulate launching of commercial satellite technology and commercial space travel
Airports	ARP	Planning and development of a safe and efficient airport system
Region and Centers	ARC	Business services to internal and external customers
Research and Acquisitions	ARA	Aviation research, competitive sourcing, procurement, navigation and surveillance systems, and air traffic systems development
Regulation and Certification	O-AVR	Medical certification of airmen, aerospace medicine, and accident investigation
Aircraft Certification	AIR	Airworthiness of aeronautical products
Flight Standards	AFS	Certification and examination of pilots and oversight of aircraft maintenance
Air Traffic Services	O-ATS	Runway safety, weather policy and standards, and system capacity planning
Airway Facilities	AAF	Maintenance of air traffic facilities and navigational equipment
Air Traffic	AAT	Safe and expeditious control of air traffic in the national airspace system from takeoff to landing

*Note: The EAS was distributed prior to the creation of the Air Traffic Organization.

The FAA was prompted by Congressional direction to assess employee attitudes following the 1981 air traffic controllers' strike. As a result, the FAA first administered the Employee Attitude Survey (EAS) to its employees in 1984. Although the content of the survey has changed over the years, many items within core areas of interest have remained unchanged. The purpose of the survey has been to collect opinions regarding organizational issues that may affect workforce performance and quality of work life, including job and pay satisfaction, attitudes toward management, and model work environment, among others. For more information about the history of the EAS, see Thompson et al., 2000.

Organizational surveys have been used as a means of seeking employee feedback, assessing reactions to organizational changes, and identifying organizational concerns (Kraut, 1996). The relationship between organizational performance and employee attitudes has also been examined. For example, the efforts of Sears, Roebuck and Company in the 1990s to transform the company's financial slump included several interventions implemented to create an "employee-customer-profit" model. Rucci, Kirn, and Quinn (1998) reported causal linkages between employee attitudes at Sears and profit. By modeling data from 800 different stores, they found that a 5-point improvement in employee attitudes led to a 1.3-point increase in customer satisfaction, which led to a .5% increase in revenue. More recent research has suggested that the financial performance of an organization may actually influence employee attitudes.

Schnieder, Hanges, Smith, and Salvaggio (2003) found employee attitudes concerning satisfaction with pay, satisfaction with security, and overall job satisfaction were correlated with financial (return on assets) and market performance (earnings per share). In testing the direction of the relationships, however, they found a stronger indication that financial performance influenced job satisfaction, rather than vice versa. Schneider et al. (2003) also found reciprocal relationships between satisfaction with pay and financial performance; that is, both influenced each other over time.

While market performance and financial return are not the outcomes of interest within the FAA, many of the organizations within the FAA use the results of the EAS to measure their progress regarding action plans established as organizational performance indicators. The purpose of this paper was to outline some of the major results of the EAS 2003 for the FAA overall, draw comparisons with other government survey results and earlier EAS data, and provide a context within which to interpret the results.

Method

The most recent administration of the EAS was a census of FAA employees in September 2003. Agency-appointed LOB/MO Points of Contact (POCs) and other survey stakeholders contributed to the survey design and content. This allowed the survey to reflect issues of interest from throughout the agency at the time of development, while maintaining core historical items. The survey was coordinated with union representatives and submitted to the FAA's institutional review board; employee participation was voluntary and anonymous.

Survey Distribution

During September 2003, approximately 48,900 surveys were mailed to all FAA employees at their work addresses. Reminder postcards were also sent to all employees. By December 2003, over 22,800 (47% of total) surveys had been returned. Of the returned surveys, 22,720 (46%) were considered “valid” (i.e., having a response to at least one non-demographic item). Table 2 presents the response rate within each organization as well as the proportion of total returned surveys for each organization.

Table 2
Response Rates by Organization

Organization Acronym	Number of Respondents	Response Rate within Organization	Percentage of Total Returned
AOA	967	52%	4%
AST	36	65%	<1%
ARP	329	69%	1%
ARC	1,262	52%	6%
ARA	928	51%	4%
O-AVR	312	60%	1%
AIR	715	62%	3%
AFS	3,040	65%	13%
O-ATS	267	46%	1%
AAF	6,058	52%	27%
AAT	8,731	37%	38%
No LOB/MO	2		<1%
FAA Overall	22,720	46%	100%

Sample Demographics

Of the 22,720 respondents, 74% were male, and 26% were female. This was consistent with the FAA’s Central Personnel Management Information System data showing that 75% of the FAA population was male and 25% was female. FAA tenure was also fairly similar between the respondent sample and the FAA population (Figure 1). With respect to job role, managers and executives represented 2% of the FAA workforce and were therefore slightly over-represented, with 6% of the survey respondents. Supervisors comprised 9% of the survey sample and 10% of the FAA workforce. Nonsupervisory employees were slightly under-represented, making up 85% of the survey respondents and 88% of the FAA workforce. Overall, survey respondents were similar to the FAA workforce in gender, tenure, and job role.

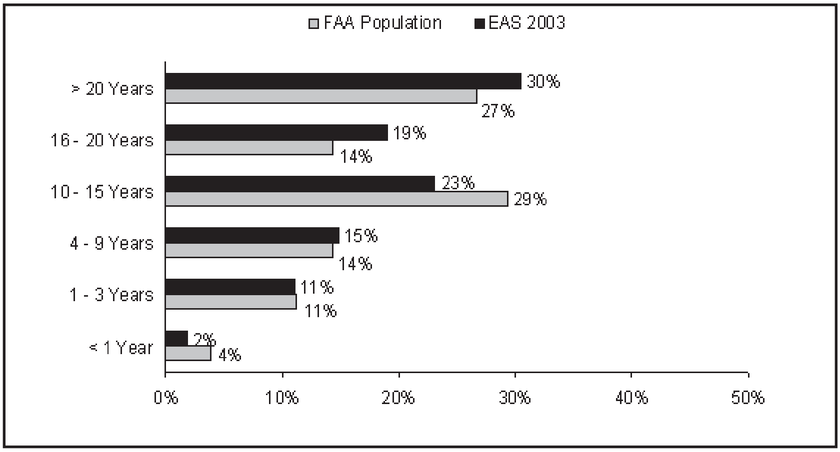


Figure 1. FAA Tenure

Survey Content

The 2003 EAS contained 129 items organized into three major sections: (1) Indicators of Satisfaction, (2) Management and Work Environment, and (3) Respondent Demographics. In addition, the survey included a section for respondents to provide comments. Each major section was subdivided into dimensions and single-items intended to measure a variety of constructs relevant to that section. The response options for the majority of items on the survey included satisfaction (i.e., very dissatisfied, somewhat dissatisfied, neither, somewhat satisfied, very satisfied), agreement (i.e., strongly disagree, disagree, neither, agree, strongly agree), and extent responses (not at all, to a limited extent, to a moderate extent, to a considerable extent, to a great extent). Demographic items, such as gender, provided categorical response options. In addition, three multiple response items encouraged participants to mark all answers that applied, such as "YES, I have been unfairly denied a career opportunity based on:" followed by a list of possible multiple response options.

Indicators of Satisfaction. Twenty items were administered to gather data on employee attitudes toward job satisfaction, supervisor satisfaction, satisfaction with compensation, satisfaction with recognition received, and organizational commitment.

Management and Work Environment. One hundred one items addressed employee attitudes toward a variety of management and work environment issues such as performance management, performance focus, resources, leadership, communication, conflict management, and model work environment (MWE). Within these broad categories were items regarding communication, recognition and rewards, supervisory fairness, employee confidence in supervisors, trust, and accountability.

Respondent Demographics. Eight demographic items were included to gather data regarding FAA tenure, present job tenure, job role, gender, region, age, education, and race/ethnicity.

Respondent Comments. Respondents were invited to provide comments at the end of the survey. A random sample of one-third of the written comments was transcribed, content coded, and quantified (King, Cruz, Jack, Thomas, & Hackworth, in press). Names and other potentially personally-identifying information were purged from comments; nonetheless, respondents were informed that their confidentiality could not be assured if the comments contained other identifying information and that transcribed comments would be subject to requests made under the Freedom of Information Act.

Data Analysis

For individual items on the EAS, frequencies and proportions were calculated for each response option. For multi-item dimensions, response distributions were calculated by counting the number of responses for each option (e.g., strongly agree, agree) across all items and then dividing by the total number of responses to all items in the dimension. Respondents only needed to answer one item in a dimension to be included in the calculation. For example, in a dimension consisting of 3 items where 20 out of 100 responses to item 1 were 5s and 30 out of 90 responses to item 2 were 5s and 20 out of 80 responses to item 3 were 5s, the proportion for response option 5 would be:

$$\frac{20 + 30 + 20}{100 + 90 + 80} = \frac{70}{270} = .259 \times 100 = 25.9$$

For negatively worded items within a dimension, the responses were reverse-scored for purposes of combining the data. For multiple response options (e.g., “mark all that apply”), frequencies for each response option were calculated.

Positive response rates were calculated by combining the frequencies for the top two response options on agreement (agree and strongly agree) and satisfaction (somewhat and very satisfied) scales. Positive response rates for extent scales were achieved by combining the frequencies for the top three response options (moderate, considerable, and great extent).

Results

Indicators of Satisfaction

Quality of Work Life. The quality of work life dimension consisted of five items addressing employees’ satisfaction with their physical working conditions, the kinds of work they do, their workgroups, organizations, and jobs overall. Results indicated that 66% of respondents reported being somewhat or very satisfied on the quality of work life dimensions, and 71% indicated that they were somewhat or very satisfied with their jobs overall. This compared with findings from two other government surveys; the Federal Human Capital Survey (FHCS[1]) reported 68%

job satisfaction (2002) and the Merit System Protection Board (MSPB[2]) survey reported 67% job satisfaction (2000). Of the major FAA organizations, Flight Standards Service (AFS) (79%) and Airway Facilities (AAF) (79%) employees were the most satisfied with their jobs overall and Air Traffic Service (AAT) (66%) employees were the least satisfied (Figure 2).

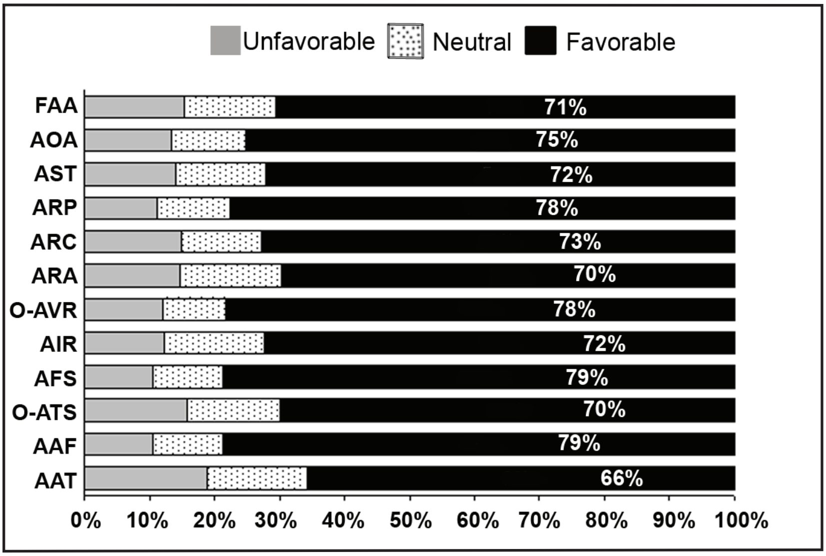


Figure 2. Job Satisfaction by LOB/MO

Satisfaction with Compensation. The satisfaction with compensation dimension consisted of three items measuring satisfaction with pay, benefits, and retirement system. Results for the FAA overall showed that FAA employees were largely satisfied with the compensation they received (65%). When asked specifically about pay, 68% of employees overall indicated that they were somewhat or very satisfied with their pay. This was a 9% increase from 59%, which was reported by FAA employees in 2000. Additionally, this compared with 64% in the FHCS survey and 49% in the MSPB survey. Within the FAA, the Associate Administrator for Research and Acquisitions (ARA) and the Associate Administrator of Airports (ARP) employees were the most satisfied with their pay (73%), while Aircraft Certification Service (AIR) employees were the least satisfied (58%; Figure 3). FAA employees, overall, appeared to be more satisfied with their pay than with their particular pay systems. Only 51% of FAA employees responded that they were somewhat or very satisfied with their pay system. A larger proportion of those on the Executive (64%) and Air Traffic pay plans (62%) were satisfied with their pay system than were those on the General Schedule (56%) and other pay plans (47%). Employees on Core Compensation (CC) were the least satisfied (38%) with their pay plan. This is important given that CC was introduced to address concerns over performance and pay issues.

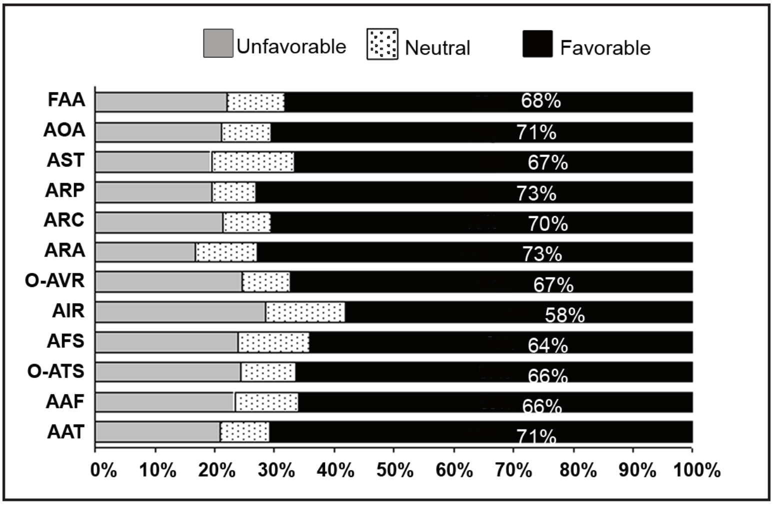


Figure 3. Satisfaction with Pay

Organizational Commitment. The organizational commitment dimension, consisting of five items, indicated that most FAA employees (81%) were committed to the FAA to a moderate, considerable, or great extent. Furthermore, 55% of respondents indicated that they did not intend to leave the FAA within the next 5 years. Of those who did intend to leave, 68% were planning to retire, and 28% planned to seek other employment.

Satisfaction with Supervisors and Recognition. The results of two single-item indicators of employee satisfaction were mixed. Most respondents were somewhat or very satisfied with their immediate supervisors (61%); however, only 38% of respondents were satisfied with the recognition they received for doing a good job. These data were consistent with what was reported in 2000. Other sources have found similar results regarding satisfaction with recognition: the FHCS survey reported 46% and 37% was reported in the MSPB survey.

Management and Work Environment

Performance Expectations. Employees indicated moderate approval concerning their understanding of performance expectations within their job (44% positive). In particular, 45% agreed or strongly agreed that communications with supervisors helped clarify what is expected in the job, and 44% agreed or strongly agreed that they were clear about how “good performance” was defined. Fewer respondents (34%) indicated that their organization had clearly communicated the connection between their individual performance goals and their organization’s performance goals, but more than half of respondents (54%) indicated that their most recent performance rating was an accurate reflection of their performance.

Job-Related Communication. Employee feedback regarding job-related communication received a 46% positive response rating. This represented a 7%

increase from 39% agreement in 2000. While more than half (58%) of respondents agreed or strongly agreed that they are encouraged to share information to get the job done, fewer indicated that policies affecting their work are communicated adequately (42%), guidance on procedures for doing their work is communicated adequately (41%), and management ensures that information needed to do their job is readily available (41%).

Recognition and Rewards. Employees were dissatisfied with recognition and rewards within the FAA (27% positive). In particular, only 20% of respondents agreed or strongly agreed that promotions are given to those who are well qualified, while 24% agreed or strongly agreed that recognition and rewards are based on merit. Twenty-nine percent of respondents indicated that people in their organization get the credit they deserve for the work they do, while 35% indicated that it's pretty common to hear "job well done" within their organization.

Accountability and Corrective Actions. Several items addressed accountability and corrective actions. For both supervisory and nonsupervisory employees, 38% of respondents agreed or strongly agreed that employees were held accountable for achieving important agency goals. However, only 21% agreed or strongly agreed that corrective actions were taken to deal with nonsupervisory poor performers. This was lower than both the FHCS and MSPB surveys, which had results of 27% and 26%, respectively. When supervisors and managers were the target of consideration, only 15% of respondents agreed or strongly agreed that corrective actions were taken with poor performers.

Conflict Management. Overall, 38% of respondents indicated that they had experienced work-related conflict to a moderate, considerable, or great extent, and 22% of respondents agreed or strongly agreed that conflicts and differences in their organization were brought out and managed, rather than avoided or worked around.

Model Work Environment. Employees were asked to rate the extent to which the FAA had created a working environment that did not tolerate discrimination, provided employees with developmental opportunities, and allowed employees to contribute to their organization's mission. Employees indicated promising results with 66% indicating a positive response for the dimension. This is a 5% increase from 61% for the FAA overall in 2000.

Communication Climate. The communication climate dimension, consisting of three items, measured fear of retaliation and the openness of the communication environment. The results were mixed with a 33% positive response rate. Specifically, 40% of respondents indicated that they were encouraged to express their concerns openly, while on the negatively worded items, 52% indicated that some employees may be hesitant to speak up for fear of retaliation, and 45% believed that it is generally safer to say that you agree with management even when you do not.

Trust. While trust is reflected in some of the previous items, particularly with regard to communication climate, several items on the EAS addressed trust directly. The majority of employees agreed or strongly agreed that they trust their co-workers (62%) and their immediate supervisors (56%). Fewer respondents (42%), however, agreed or strongly agreed that supervisors where they work trust employees. Overall, 23% of respondents expressed trust in FAA management. Although quite low, this represented an increase of 6% from 17% trust in management for the FAA overall in 2000.

The level of trust employees expressed provided an interesting look at the effect of supervisory/nonsupervisory role (Figure 4). As the job role of the respondent was closer to the top management level (i.e., from nonsupervisory employees to supervisors, managers, and executives), the percentage of employees expressing trust in supervisors and management also increased. Nonsupervisory employees had the lowest rates of trust in immediate supervisors (53%) and FAA management (19%).

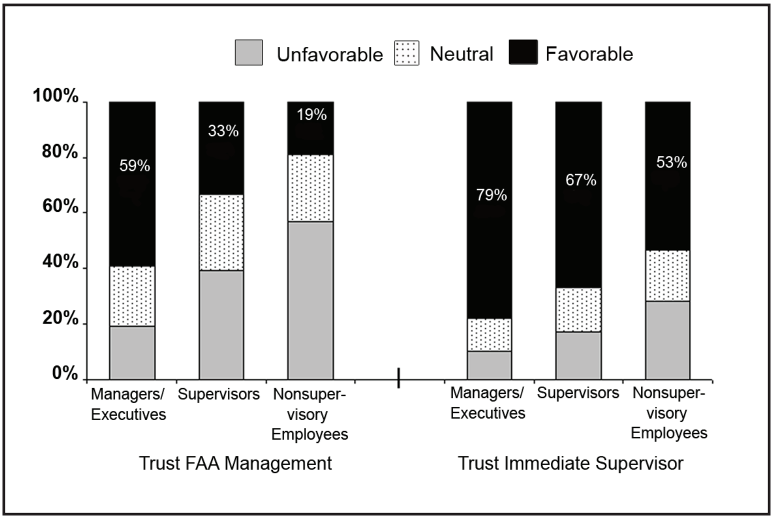


Figure 4. Trust in FAA Management and Supervisors by Job Role

In spite of the results regarding trust in supervisors and management, employees' confidence in their supervisors remained consistent with 59% agreement on this dimension in both 2000 and 2003. Perceived supervisory fairness also remained consistent, with 54% agreement on this dimension in both 2000 and 2003. Finally, 56% of employees agreed that supervisors facilitated problem solving.

Discussion

In general, the FAA workforce reported a high level of commitment and job satisfaction. However, respondents to the 2003 EAS generally did not believe that poor performers (employees or management) were held accountable. This sentiment is not unique to the FAA and has been a frequent topic in business literature. In a *GovExec* article, Shoop (2004) suggested that accountability has become meaningless, particularly in the government. Central to the success of performance management, employees must believe it is a fair process (National Academy of Public Administration, May 2004).

Attitudes toward pay and pay systems are particularly relevant to perceptions of fairness in performance management. For the most part, FAA employees were satisfied with their pay; however, feelings toward their pay system were less positive for some, particularly those on the Core Compensation (CC) pay system. These results are troubling given that the FAA is transitioning to the CC system in an effort to become more performance-based. Currently, two types of annual increases (i.e. superior contribution increase [SCI] and/or organizational success increase [OSI]) are available to employees within the CC system. Employees have the possibility of earning an SCI of 1.8% or 0.6%. In addition, an OSI is available if the organization as a whole meets a percentage of the goals for the year as determined by the FAA administrator. Employees may earn both types of annual increase or only an OSI. Based upon their 2003 performance, employees under CC were eligible to receive as much as a 4.95% increase (not including locality pay).

Reasons for the lack of satisfaction with the CC system may include perceived low pay-off for high-performing employees, increased workload for managers and supervisors as a result of the CC system, and a feeling of unfairness due to the low number of employees under CC relative to the population of the FAA.

In addition to low levels of satisfaction with the CC system and accountability of employees and management, results also revealed very low levels of trust in FAA management, particularly among nonsupervisory employees and supervisors. Not surprisingly, trust in FAA management was higher among managers and executives. These findings compared with the Watson Wyatt WorkUSA (2002) survey of nearly 13,000 private sector employees, which found that only 39% of employees trusted their senior management. Based on these data, it would appear that the level of trust in management within the FAA is significantly lower than in the private sector. Trust in management has been shown to influence shareholder return rates in the private sector and the extent to which employees feel committed to the organization, the extent of cynicism about change, and employee intent to leave in public sector organizations (Albrecht & Travaglione, 2003).

These areas will need to be reviewed by upper management to understand how best to link accountability, performance, and pay. Follow-up discussions with employees around these issues could afford FAA policy makers with important feedback necessary to strategize interventions or modifications.

Successful survey programs have clear, responsive, and well-defined action plans. Further, managers should be provided with responsibilities for follow-up (Gilbert, Slavney, & Tong, 2003). Employees should be informed of the area(s) chosen for intervention and provided with information regarding the implementation and status of action plans. Engaging employees invites them to affect the organization's future (Burke, Coruzzi, & Church, 1996). When employees do not perceive responsive action to the results of an organizational survey, it can undermine their willingness to participate in the survey process. Employees may withdraw from future participation and, further, lose trust in management due to the perceived lack of follow-through. Recently, the FAA administrator addressed the EAS 2003 results by designing a corporate action plan. Several areas were identified as needing improvement: conflict management and resolution, lack of clarity regarding performance expectations, recognition and rewards, and communication (http://employees.faa.gov/news/admin_message/index.cfm?admin_message=dsp_admin_arch_062504.cfm). To address conflict management, an Early Dispute Resolution Center is to be created at headquarters. As a follow-up for communication, an independent firm conducted an assessment of the agency's internal communication procedures. With the assistance of FAA executives, a number of actions aimed at improving FAA's communication were developed in response to the findings (http://employees.faa.gov/library/publications/comms_rep/media/commplan.pdf). The full architecture and impact of these programs are yet unknown. Nonetheless, more than 22,700 FAA employees provided their input and feedback, both positive and negative. Communicating actions and achievements resulting from the EAS 2003 is critical to maintain or improve upon employee satisfaction and ensure participation in future EAS administrations.

References

Albrecht, S., & Travaglione, A. (2003). Trust in public-sector senior management. *International Journal of Human Resource Management, 14* (1), 76-92.

Burke, W.W., Coruzzi, C.A., & Church, A.H. (1996). The organizational survey as an intervention for change. In A.I. Kraut (Ed.), *Organizational surveys: Tools for assessment and change* (pp. 41-66). San Francisco, CA: Jossey-Bass.

Department of Transportation. (2003). *Total employment by administration*. Retrieved June 7, 2004, from http://dothr.ost.dot.gov/Workforce_Information/workforce_information.html.

Federal Human Capital Survey. (2002). *Results from the Federal human capital survey.* Retrieved June 8, 2004, from http://www.fhcs.opm.gov.

Gilbert, P., Slavney, D., & Tong, D. (2003). 10 best practices for employee surveys. *Workforce Management On-line Magazine.* Retrieved July 12, 2004, from http://www.workforce.com.

King, S. J., Cruz, C., Jack, D. G., Thomas, S., & Hackworth, C. (In press). *2003 Employee survey analysis of employee comments.* (DOT/FAA/AM-04/in press). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Kraut, A.I. (1996). An overview of organizational surveys. In A.I. Kraut (Ed.), *Organizational surveys: Tools for assessment and change* (pp. 1-14). San Francisco, CA: Jossey-Bass.

Merit Systems Protection Board. *The federal workforce for the 21st century: Results of the merit principles survey 2000*. Retrieved June 15, 2004, from http://www.mspb.gov/studies/ mps_2000/mps_2000.htm.

National Academy of Public Administration (2004, May). *Recommending performance-based federal pay*. Retrieved July 14, 2004, from http://www.napawash.org/Pubs/Broadbanding5-04.pdf.
Rucci, A.J., Kirn, S.P., & Quinn, R.T. (1998). The employee-customer-profit chain at Sears. *Harvard Business Review*, *76*, 83-97.
Schneider, B., Hanges, P.J., Smith, B., & Salvaggio, A.N. (2003). Which comes first: Employee attitudes or organizational financial and market performance? *Journal of Applied Psychology*, *88*, 836-51.
Shoop, T. (2004, May 17). Accountability lite. *GovExec*. Retrieved May 20, 2004, from http://www.govexec.com/dailyfed/0504/051704ol.htm.
Thompson, R., Hilton, T., Twohig, P., Pagnini, C., Park, H., King, S.J., Malone, M., Thompson, D., & Thompson, J. (2000, March). *Results of the 1997 employee attitude survey.* (Memorandum Report). Oklahoma City, OK: FAA Civil Aerospace Medical Institute Human Resources Research Division.
Watson Wyatt Worldwide (2002). *WorkUSA 2002 weathering the storm, A study of employee attitudes and opinions*. Washington, DC.

Footnotes

[1]The Office of Personnel Management administered the Federal Human Capital Survey (FHCS) in 2002 to a random sample of government employees across 24 agencies. Over 100,000 employees responded.

[2]In 2000, the Merit System Protection Board (MSPB) randomly selected government employees across 23 agencies. Nearly 7,000 employees responded.

International Journal of Applied Aviation Studies, 5, Number 2

Understanding Consumer Preference–Findings from the Airline Survey

Brent D. Bowen
University of Nebraska at Omaha
6001 Dodge Street, AH 422
Omaha, NE 68182
402-554-3424
bbowen@mail.unomaha.edu

Nanette Scarpellini-Metz
1648 Splinter Rock Way
North Las Vegas, NV 89031
702-649-0103
nscarpel@unoamaha.edu
and

Dean E. Headley
Wichita State University
W. Frank Barton School of Business
304 Clinton Hall
Wichita, KS 67260-0084
headley@aqr.aero

Abstract

The Airline Survey of frequent travelers provided insight into consumer preference in four distinctive areas of airline travel. This study was intended to offer a better understanding of travelers' behavior through ranking of specific options offered by the airlines as well as their response to air transportation issues. The first area explored the basis and range of quality highlighting the perceptions of frequent travelers. This section correlated with the Airline Quality Rating (AQR) which focuses on airline performance features significant to air travel consumers. The second area provided insight into what factors go into selecting a specific airline. This included the level of importance participants assigned to key concerns. The third area focused on consumer complaints and tied into the United States Department of Transportation Air Travel Consumer Report for added validity. Complaints typically centered on the issues of handling cancellations or delays. Rude treatment by airline em-

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

ployees and loss of luggage also were significant problems. The fourth area explored the impact on airline travel from the terrorist attacks of September 11, 2001 (9/11). This included the number, if any, of business and or pleasure trips that were canceled in response to 9/11. The survey results indicated the majority of respondents did not cancel any trips. Of those that did cancel, typically only one trip was affected. Overall, the survey results provided an indication of consumer behavior and response to the changing air transportation environment within the demographic of frequent business and leisure travelers.

Understanding Consumer Preference—Findings from the Airline Survey

The Airline Survey of frequent flyers was conducted by the Aviation Institute at the University of Nebraska at Omaha and Wichita State University. The survey instrument was constructed as a partial revalidation of the Airline Quality Rating (AQR) developed in 1991. The researchers wanted to find out what were currently the most important issues to frequent travelers. The results provided an indication of consumer behavior within the demographic of frequent business and leisure travelers. The responses were based on a survey of 120 items. Of the 2000 surveys distributed, 766 were returned with usable responses.

The current AQR report was based on a summary of monthly AQR scores from the previous year. These scores centered on 15 components divided into four major categories that brought together airline performance features significant to air travel consumers (Bowen & Headley, 2004). The scores rank all United States (US) airlines that carry a domestic passenger volume of at least 1% for that year. The AQR uses a "system of weighted averages and monthly performance data in the areas of on-time arrivals, involuntary denied boardings, mishandled baggage, and a combination of 12 customer complaint categories" (para. 2). The AQR was designed to provide a weighted average of elements determined to be important to consumers of airline services. To be included on the rating scale, the elements were not only relevant to the consumer, but also measurable. The element had to be available for each airline from a published data source. The weights were verified by using an expert panel of the airline industry to rate the level of importance to the consumer of each element in terms of airline quality. "Weights reflect importance of the criteria in consumer decision-making, while signs [positive or negative] reflect the direction of impact that the criteria should have on the consumer's rating of airline quality" (para. 5). The combined result was a single interval scaled value that can be used to compare all airlines regardless of time period. To maintain its position as an industry standard, this study was designed to revalidate the AQR and to reveal any discrepancies with the AQR based on industry and societal changes. This study was the first in a series of papers to analyze and assess the data.

Literature Review

The aviation industry experienced many changes since its inception in the early part of the 20th century. Leading up to the terrorist attacks of 9/11, the industry was trying to come to terms with addressing consumer preferences and demands in an environment of rising costs and increased security concerns. While it was doubtful any industry could be prepared for such an event, the industry is now addressing its current condition. To do so, the aviation industry must acknowledge its accomplishments and failures and build a plan for recognizing current and future needs. The following literature laid the groundwork for preparing the Airline Survey and analyzing its results.

Airline Consumer Preferences

When examining consumer preferences, the airline industry has been left in a vulnerable position. The passenger focused on two primary areas of airline service: schedule and price (Headley & Bowen, 1997). Other issues, while important, maintained a secondary role. "Service quality perceptions are often defined as the difference between what a consumer expects and what they actually perceive as outcome" (Headley & Bowen, 1997, p. 55). With the increasing number of competitive options in the 1990s, a viable market strategy was difficult to maintain (Ott, 1998). The airlines were realigning themselves to address these preferences in order to maintain customer retention. Consumer loyalty was no guarantee. Lower prices and/or preferred flight times held as much influence as customer satisfaction when a purchase was made. The Airline Quality Rating findings, since its inception in 1991, have revealed problem areas in airline service and provided insight into possible solutions for both the airlines and aviation authorities attempting to promote air transportation quality to current and potential consumers (Bowen, Headley & Lu, 2003).

Customer loyalty to a specific airline typically was based on a higher perceived quality of one airline over its competitors (Fick & Ritchie, 1991). In contrast, fewer differences among competitors decreased the probability of forming brand loyalty. If the airline delivered satisfactory service, the consumer's loyalty was reinforced to the specific provider (Bitner, 1990). When there is little to differentiate the airlines, consumers' perception of service quality is a significant issue (Rhoades & Waguespack, 2000). Service quality is more than providing reliable service at the basic level, it also must "involve understanding customer expectations and perceptions and then meeting or exceeding them" (p. 62). In these circumstances, the airlines and their employees listened to the customer and responded to service problems. In a Frequent Flyer Survey (1997), basic service and service design controlled the three key factors to airline satisfaction: on-time performance, airline schedule, and flight check-in. These factors continue to be primary areas of concern.

A survey conducted by a travel research firm found that price was the primary subject when purchasing a ticket followed by the preferred airline and the best connections (Consumer Reports, 2000). Price was a major concern for leisure and business travelers. In the early 1990s, business travelers, individuals traveling at the expense of their employers, made up about one-third of airline passengers, but accounted for about two-thirds of the passenger revenues (Stephenson & Bender, 1995). The remaining two-thirds of the passengers were leisure or personal travelers. They found, at the same time, that the Air Transport Association of America noted a gradual decline in the number of business travelers. The high cost of airline travel for business travelers was viewed as a major deterrent. To avoid paying for the premium ticket, businesses were increasing the use of teleconferencing and automobile travel for the shorter trips from 200 to 400 miles and less (Stephenson & Bender, 1995).

Consumer emphasis on price and flight schedule often outweighed other factors when purchasing services (Ott, 1998). The Air Traveler Consumer Report, published quarterly by the Office of Aviation Enforcement and Proceedings under the US Department of Transportation, provides details on the major US carriers (2005). The report relates information regarding on-time percentage, mishandled baggage reports, passengers denied boarding, and consumer complaints by category (Rhoades & Waguespack, 2000). Matters of brand identification held little power to many consumers. Even consumers with a preferred airline selected another carrier to save money or get an ideal departure time. On short flights, the amount of product differentiation, through advertising, customer service or brand identification, was not significant enough to outweigh the price and time factors (Ott, 1998). As airlines carried more passengers in more cramped conditions, the distinction between competitors narrowed.

Regional jet service received a mixed response from travelers who often felt like second class citizens left out in the cold (The Boyd Group, 2004c). Since jet ways do not always coordinate with regional jet configurations, passengers frequently must leave the comfort of the air terminal to board and deplane. This means waiting in the cold and even rain or snow to retrieve oversized carry-on baggage that would typically fit on a standard jet's overhead compartment. This becomes more than a minor inconvenience for many travelers who opt out of this form of travel for the comfort of a major carrier. While too many empty seats drain profits when mainline aircraft fly smaller demand routes, overbooked regional jets create cramped and unhappy passengers (Embraer, 2004). As regional markets expand, Embraer says passenger demand is moving toward 70 to 110 seat segment. This equipment range is just now being developed with Embraer's new line of commercial jets. With regional jets gaining in prominence, the length of the flight also grows. Therefore, there is greater need for enhanced cabin comfort and other passenger facilities such as the baggage and cargo services associated with those on a mainline aircraft. By using an aircraft sized for the route, the airline minimizes excess seat capacity as well as reduces wasted fuel and unnecessary weight-related charges (Embraer, 2004)

On the other end of the spectrum, micro business jets seating four to six passengers are gaining appeal as the price of business class rises. If a company is sending three of four representatives, it may be cheaper to charter an air taxi (The Boyd Group, 2004b). The company will also save time and bypass congested airports.

Industry trends change customer relationships. With the increase and expansion of internet use, consumer behavior now reflects a more independent and savvy attitude (The Boyd Group, 2004a). Travelers can be their own travel agents and expect the airlines to match the best fare. This also contributes to the decline of loyalty programs for shaping the consumers' buying decisions. With such easy access to fares, trips are able to be more spontaneous than in the past. The

Boyd Group suggests the pricing will become simplified by 2005, as airlines such as Alaska Airlines, US Airways, and Delta launch new nationwide pricing plans. Computer Reservation Systems will fade as internet fares become more common and accepted.

Status of the Aviation Industry Prior to 9/11

The aviation industry has been subjected to various conflicts shaped by internal and external forces. Between 1993 and 2002, three Blue Ribbon Commissions were formed by presidential decree to study the aerospace industry. The first two, the National Commission to Ensure a Strong and Competitive Airline Industry of 1993 and the White House Commission on Aviation Safety and Security of 1997, had little impact on improving the security of the nation's airports (Scarpellini-Metz & Bowen, 2004). Even though security was a central issue of the 1997 report, also known as the Gore Report, little active attention was paid to the recommendations for improving security measures. Leading into 9/11, consumer preference seemed to control the way, with an emphasis on low fares and improved access (Cobb & Primo, 2003). Consumers were favoring new lower costs airlines, such as Southwest, that offered inexpensive, no-frills flights.

The airline industry is extremely reactive to the variable cycles of the economy. According to Ghobrial and Irvin (2004), this industry "is perhaps the first to feel the negative affects of a weak economy, and could be the last to feel the positive affects of a rising economy" (p. 70). Air carriers remained under the control of the Federal Government from the 1920s until the passage of the Airline Deregulation Act of 1978 (Kane, 2003). "The phasing out of government regulation of domestic airline operations has provided convincing evidence…[the] transportation consumer can benefit through a wider choice of services and prices" (Kane, 2003, p. 307). The airline deregulation led to major structural and management changes in US airlines (Ghobrial & Irvin, 2004).

Prior to the terrorist attacks of 9/11, the US air transportation industry was suffering from increasing labor costs and decreasing travel demand (Chui & Bowen, 2004). "High fuel prices, rising insurance costs, and the added costs of ensuring security have all contributed to the troubled condition of the aviation industry" (Ghobrial & Irvin, 2004, p. 68). The events of 9/11 dealt a crushing blow to an already vulnerable aviation industry. By November of 2001, the US economy officially entered its 10th recession since World War II (Rupp, Holmes & DeSimone, 2003). As of January 2001, business travel dropped significantly reflecting the US economic decline. These same events also forced the global aviation industry into an economic crisis (Bisignani, 2002). Air travel demand dropped radically in 2002 (Office of Aviation Policy & Plans [OAPP], 2003). "Globally, 400,000 aviation-related jobs were eliminated and "airlines lost 18 billion dollars in 2001—12 billion on international scheduled services and 6 billion on domestic services" (Bisignani, 2002, para. 3). Passenger service levels rise and fall due to a number of conditions including competition, economic trends, and terrorism (Berry, Bhadra, Gentry, & Nelsoni, 2004). Not all airlines were able to respond successfully to these challenges. As of 2003, the FAA projected a world passengers' traffic average to maintain an annual growth of 5.2 percent from 2004 to 2014 to total 2.0 billion (OAPP, 2003). While most of the world has recovered and shown gains since 2001, for North American carriers, the air traffic in the first quarter of 2004 was still 2.8% below 2001 figures (Bisignani, 2004, line 34). Based on historical trends of US economic strength, the economy will return to normal by 2005 (Chui & Bowen, 2004, p. 3).

When the US aviation industry seemed isolated from attacks overseas, issues of safety and security were often overshadowed. As the events faded from the news media, so did the support of security measures. Prior to 9/11, the US government focused on creating a safer flying environment with the Safer Skies initiative launched by the Clinton Administration in 1998 (Oster, Strong, & Zorn, 2000/2001). This safety program called for an 80% reduction in fatal accident rates by 2007. At the time, the focus was on three areas: lessons learned from previous accidents, investigate cause of accidents, and a collection of issues such as passenger interference with crew, seat belt use, carry-on baggage, and child safety restraints (Oster, et al., 2000/2001). Ground congestion was also seen as a major problem, which included such threats as runway incursions and collisions with ground support vehicles, other aircraft, and pedestrians. The safety focus was on improving accident investigation technology through flight deck video recorders as well as increasing automation both on the flight deck and with air traffic control. Reducing the risk of terrorism was viewed negatively since it would be costly, inconvenience passengers, and add to the delays system-wide. Many of the possible measures were met with resistance due to their possible violation of passenger privacy (Oster, et al., 2000/2001). For most countries, as well as the US, safety regulation is a government function. At that time, the concept of safety centered on aircraft viability not on terrorist identification.

The AQR provided a consistent model for assessing the airline industry that can be measured and compared over time. In a study of the AQR for years 1999 through 2003, the researchers found a decline between 1999 and 2000, but improvement between 2001 and 2002 even with 9/11 (Bowen, Headley & Lu, 2003). While the terrorist attacks had a negative effect on the aviation industry overall, the level of service quality improved. “Overall air service was observed to be at the lowest in 2000” (p. 16). The worst service typically occurred during the winter and summer months: December, January, June, July, and August. However, this improvement was threatened by regressive actions by the airlines and the FAA. As small-capacity aircraft fill airport slots and airspace, over-scheduling during peak times led to increased delays at major airports while non-hub airports suffered from the decrease in revenue passengers (Bowen, Headley & Lu, 2003). Between airport security procedures, aviation safety regulations, and an antiquated National Airspace System, the needs of the traveling public are being neglected. A collaborative effort between the government and airlines should address the passenger needs now and in the future (Bowen, Headley & Lu, 2003).

Method

The Airline Survey was developed based on multiple stages of assessment and evaluation. The core of the survey was constructed to correlate with the major categories of the AQR and their components. Due to the volatile nature of the airline industry, particularly after the terrorist attacks of September 11, 2002, the mass media was analyzed to define issues now facing frequent travelers. During this period, news about air travel saturated the news and people were more aware of the issues. The survey then went through two panels of experts in the areas of research and frequent travel before the final version was prepared for distribution. The responses were based on a survey of 120 items.

The Airline Survey was distributed in 2002 to 2000 individuals. The distribution list was based on a random sample of airline frequent flyer program members supplied by a marketing firm. The surveys were collected and went through preliminary analysis in 2003. Of the 2000 surveys distributed, 766 were returned with usable responses. The return rate was 38.3%.

Research Questions

The airline survey was constructed to address the study's research questions. The researchers wanted to find out if the respondents would revalidate the Airline Quality Rating (AQR) factors with their responses as to what is the most important to frequent travelers.

Primary research question:

- Does the data suggest that the factors established in the AQR are still current and valid?

Secondary research questions:

- Will the AQR be revalidated?
- What is the sense of the frequent travelers of the AQR's subtopic areas?

Participants

Participants in this study were selected from a random sample of multiple airlines' frequent flyer programs. The sample was provided by a marketing firm based in Wichita, Kansas. In total, 2000 individuals were selected to receive the study's survey. Out of the only mailing, 766 surveys were returned with usable responses. No attempts were made to follow-up with nonrespondents to increase the return rate.

Survey Development

The survey went through a multi-stage analysis and development consisting of aviation content experts in research and frequent travelers. Individuals were termed experts based on their role in the industry and proven record through publications. In the first stage, the AQR was used to form the core area of questions. From that core, questions were constructed based on readily identifiable issues before the frequent travelers from the media. In the third stage, the instrument was distributed to an expert panel that was also a collection of frequent travelers. The panel reviewed the initial content and added to the preliminary list of questions. In the final stage, the revised survey went to a different panel of experts for validation. The second panel was a convenience sample of randomly available experts including members of the *Journal of Air Transportation's* Panel of Reviewers and Editorial Board.

The final Airline Survey was composed of 98 items presented over five pages. The items were presented in a variety of open and closed formats including Likert scale, fill-in, ranking preferences, and yes/no questions.

The surveys were sent shortly after the terrorists' attacks of September 11, 2001. During this period, news about air travel saturated the mass media and people were more aware of the issues. The mailing included a cover letter, instructions, and a reply envelope. The researchers did not follow-up with non-respondents.

Unit of Analysis

The unit of analysis for this study is based at the individual level. The respondents answered as individual entities with no identifiable connection to each other or to specific organizational affiliations. Their responses were to be representative and generalizable to the population of frequent travelers.

Validation. The survey was modified as a result of the input of acknowledged experts in the areas of aviation, research methods, and frequent travelers. "A design is internally valid if it is free from nonrandom error or bias" (Fink, 1995, p.56). Using the two panels of experts was an excellent validation opportunity since each panel was composed of individuals skilled in research methods that were also frequent travelers.

Survey Findings and Results

The survey was developed to gather information on consumer preferences in four distinct, but related areas of air travel. The data from the survey was divided to explain these four areas while acknowledging some results easily could be applied to several issues. By looking at the issues of quality, airline selection, consumer complaints, and impact of 9/11 on air travel, the data reveals useful insight into consumer preferences.

Basis and Range of Quality

The AQR establishes the quality of an airline based on four categories of importance to the traveler: on-time arrivals, involuntary denied boardings, mishandled baggage, and a grouping of 12 customer complaint categories. Various elements of the Airline Survey explore different aspects of these categories as a whole and, in some cases, individually.

One section of the survey asked respondents to rate the importance of airline performance areas to their perception of "quality" airline performance for the carriers serving the US domestic market. This part of the survey included 20 performance areas that the respondents rated on a 0 -10 Likert scale. The score of 0 *Indicated Not At All Important* and 10 was *Absolutely Essential*. The four categories of the AQR were included in these performance areas. However, only three of the AQR areas fell in the top five. The fourth category held the 12th spot.

The respondents shared a consensus on the areas of highest importance (Table 1). Having baggage arrive with the traveler, received a ranking of 9 or 10 by 87.3% of the respondents. This outscored another area of critical importance: Safety record of airline. 71.35% of frequent travelers responding ranked safety record of airline as a 9 or 10. In the third spot, 67.2% respondents indicated the need for competitive ticket prices. Handling of customer complaints and On-time arrival of flight rounded out the top five spots in terms of being absolutely essential or just below.

Table 1

Frequent Traveler Rating of Airline Performance Areas

Performance Area	% Rated 10	% Rated 9	Total % of 9 and 10	Rank
Baggage arriving with you	72.0	15.3	87.3	1
Safety record of airline	56.5	14.8	71.3	2
Competitive ticket prices	44.5	22.7	67.2	3
Handling of customer complaints	30.0	22.5	52.5	4
On-time arrival of flight	27.3	24.4	51.7	5
Courteous treatment by airline personnel	28.8	18.7	47.5	6
Leg room/seat pitch	21.9	22.1	44.0	7
Customer service	23.6	19.6	43.2	8
On-time departure of flights	22.4	20.2	42.6	9
Jet aircraft vs. Prop jet aircraft	24.6	17.5	42.1	10
Convenient schedule of flights	18.1	23.2	41.3	11
Not overbooking flights	17.7	16.9	34.6	12
Frequent flyer benefits	15.8	17.4	33.2	13
Width of seat	14.9	14.3	29.2	14
Carry-on baggage space	11.6	16.8	28.4	15
Electronic ticketing	12.3	15.3	27.6	16
Type of aircraft	9.5	12.4	21.9	17
Size of aircraft	7.9	11.4	19.3	18
Availability of upgrade seats	9.2	9.9	19.1	19
On-line check-in	8.9	8.5	17.4	20

Note: 10 = Absolutely Essential rating; 9 falls just below 10 on a 0 to 10 Likert scale.

On Time Arrivals

As an industry, the 2004 AQR report found the on-time arrival rate decreased slightly from the previous year (2005 AQR, 2005). The arrival rate went from 82.1 % in 2002 to 82% in 2003. The US Department of Transportation (OAPP) describes a flight as “on-time” when the flight is “operated within 15 minutes of the scheduled time shown in the carriers’ Computerized Reservations System” (Bowen & Headley, 2004, p. 57). The survey results found 51.7% of respondents ranked the on-time arrival of the flight as a 9 or 10 on the Likert scale. This represents a small majority of the respondents thus suggesting that while this issue is important, there are other performance areas to consider that rate higher overall.

Involuntary Denied Boardings

The area of involuntary denied boardings in the AQR is based on the description of the "number of passengers who hold confirmed reservations and are involuntarily denied boarding on a flight that is oversold" (Headley & Bowen, 1997, p. 57). In these instances, the oversold flight departs without the passengers. Involuntary denied boardings per passenger served increased in 2003 to .86 per 10,000 passengers compared to .72 per 10,000 passengers in 2002 (Bowen & Headley, 2004, p. 8). Even though overbooking may have increased, the survey found that frequent travelers do not rate this issue high when considering the quality of an airline's overall performance. Only 17.7% of respondents indicated that this was an Absolutely Essential area. This could be because frequent travelers have developed strategies to avoid being bumped. Another section of the survey found that only 7% of the 712 travelers responding (93%) had complained of being bumped recently. Whereas 14.4% of the 656 respondents (85.6%) indicated they voluntarily gave up their seats on an overbooked flight. Both numbers are nearly evenly split between business and pleasure trips.

Mishandled Baggage

The AQR category on mishandled baggage is based on the Department of Transportation's description of a mishandled bag as one in which the passenger "claims for lost, damaged, delayed, or pilfered baggage" (OAPP, p. 57). This category ranked first on the Likert scale rating detailing the perception of quality airline performance. Even though it ranks so high, 11% of all respondents indicated they had recently complained about lost or damaged baggage or carry-on policies. Whereas, 36% of the 702 respondents (91.6%) indicated that their checked baggage has been lost or mishandled temporarily in the previous 12 months. In a similar question, 2.2% of the 589 who responded said that their checked baggage had been lost and never returned. The only areas complained about more often were the manner cancellations and/or delays were handled (16.2%) and rude treatment by airline employees (11.2%). The AQR also noted an increase in mishandled baggage rates from 3.84 per 1,000 passengers in 2002 to 4.00 per 1,000 passengers in 2003 (Bowen & Headley, 2004, p. 8). The survey found that the value of the contents and/or bag were the prime consideration when travelers were deciding whether to carry-on or check the baggage.

Airline Selection Factors

In this section of the survey, participants were asked to assign points to 10 areas that pertained to their selection of a particular airline for travel. Among the 10 areas, participants were asked to assign points from 0 to 100 that reflected the importance of this area when selecting an airline. If desired, all 100 points could be assigned to one area. The 10 areas were selected based on mass media coverage of airline selection and expert opinion. When analyzing the results, the researchers looked at the total number of points assigned to each area. Out of a possible 76,600 points with a n of 766, 97.3% of the points were assigned. The remaining 2.7% or 2100 points were left unassigned without any explanation. Perhaps the participants did not fully understand this section of the survey or there was difficulty in determining when the individual had assigned a total of 100 points. Survey responses were checked to make sure a respondent did not assign more points than allotted.

The ranking of each of the 10 areas is reflected in Table 2. Considering the media coverage of air transportation safety following 9/11, the low ranking 6th place of security procedures when selecting an airline may be surprising. With only 5324 points or 7.1% of total assigned to this area, the travelers participating in this survey appear to indicate an overall trust in the security of air travel.

Table 2

Airline selection factors ranking of points assigned by respondents

	Grouping by Points Assigned					
Airline Selection Factor	**0**	**1-10**	**11-25**	**26-50**	**51-75**	**76 -100**
Price of ticket	8	154	224	228	48	32
Airline goes where needed	11	267	265	120	52	5
Convenient flight schedule	31	379	151	32	0	2
Frequent Flyer program consideration	70	312	111	43	4	3
Airline safety record	39	301	140	47	21	2
Reliable past performance	51	343	71	4	1	0
Overall quality perceptions	80	317	27	5	0	1
Aircraft cabin comfort	59	321	63	20	1	2
Airline safety procedures	70	260	110	26	3	1
Type of aircraft	140	248	27	4	0	0

The most important factor, with nearly double the points of any other area, is price of the ticket. Participant assigned 28.7% of the points to price of ticket. The next closest area was goes where needed with 17.1% of the points. The third highest area was the safety record of the specific airline with 7434 points or 9.9% of the total. Safety and security may have fallen to its lower position because the issue was divided into these two sub-areas. If they are combined, they would account for 17.1% of the points—equal with the goes where needed area, but still far behind the price of the ticket. Type of aircraft was the least important area and received only 2211 or 3% of the points.

Ranking of Customer Complaints

Nearly a third of Airline Survey respondents (30%) rated the handling of customer complaints to be Absolutely Essential (Table 1). The issue of customer complaints was examined in various sections of the survey including time to respond to complaint and complaint resolution. The AQR includes 12 specific subcategories to compose the primary customer complaint section. These complaints are flight problems; oversales; reservations, ticketing, and boarding; fares; refunds; baggage; customer service; disability; advertising; discrimination; animals; and other. Consumer complaint rates actually declined slightly from 1.22 per 100,000 passengers in 2002 to .67 per 100,000 passengers in 2003 (Bowen & Headley, 2004, p. 8). In another section, travelers were asked to identify the nature of their most recent complaint from a series of issues that closely paralleled the AQR complaint subcategories (Table 3, Figure 1).

Table 3
Percentage of Respondents' Nature of Most Recent Complaint

Nature of Recent Complaint	% of Respondents	Complaint Rank	AQR Complaint
Manner cancellations and/or delays were handled	16.2	1	Flight problems
Being bumped from a flight	2.9	8	Oversales
Reservation, ticketing or boarding problems	7.3	5	Reservation, ticketing, and boarding
Incorrect information about fares	.9	11	Fares
Getting refunds/adjustments in fares	2.2	9	Refunds
Lost or damaged baggage or carry-on policies	11.0	3	Baggage
Customer service	9.0	4	Customer service
Treatment of passenger with disability	1.4	10	Disability
Unfair, misleading or offensive advertising	.5	12	Advertising
Rude treatment by airline employees	11.2	2	Discrimination
Use of frequent flyer benefits	4.0	7	
Other	6.3	6	Other

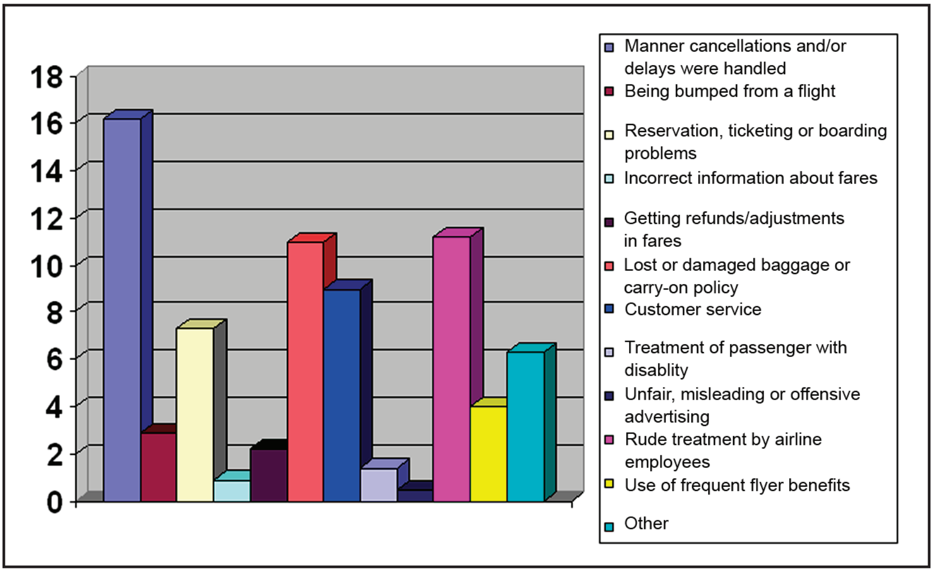


Figure 1. Percentage of Respondents' Nature of Most Recent Complaint

The AQR also includes a subcategory complaint covering animal-related complaints. These types of complaints are based on the loss, injury, or death of an animal while being transported by an air carrier (Bowen & Headley, 2004, p. 59).

In the area of *Other* in the Airline Survey, the level of unprofessional and potentially unsafe behavior was described by 14 respondents. This included two people who indicated they felt sexually harassed and groped by airline personnel as well as one respondent's whose pilot consumed an alcoholic beverage prior to the flight. Damaged and faulty equipment was another area of concern for seven respondents. This incorporates such areas as a dilapidated aircraft, to broken seats in first class, and an inoperative entertainment system.

The issue of complaints brought in many different concerns. Of the respondents who answered the question, just over half (50.4%) *personally experienced a situation with airline travel that you felt warranted a formal complaint to be registered with an airline*. Of those who experienced the situation, 71.1% indicated they had filed a complaint about the airline service with an airline, a government agency, or any other group. Of those that complained, 59.6% had their complaints responded to in less than 30 days. However, 27.1% never received a response regarding their complaint. Over half of the complaints (53.5%) were not resolved to the respondents' satisfaction.

These findings suggest the Customer Service Plans released by the Air Transport Association members in September 1999 have failed to be fully implemented (Office of Aviation Enforcement, 2005). The plans included such sections as

- notifying passengers of known flight delays and cancellations
- meeting customers' essential needs during long on-aircraft delays
- allowing reservations to be held or tickets to be refunded within 24 hours of purchase
- being more responsive to customer complaints (para. 2)

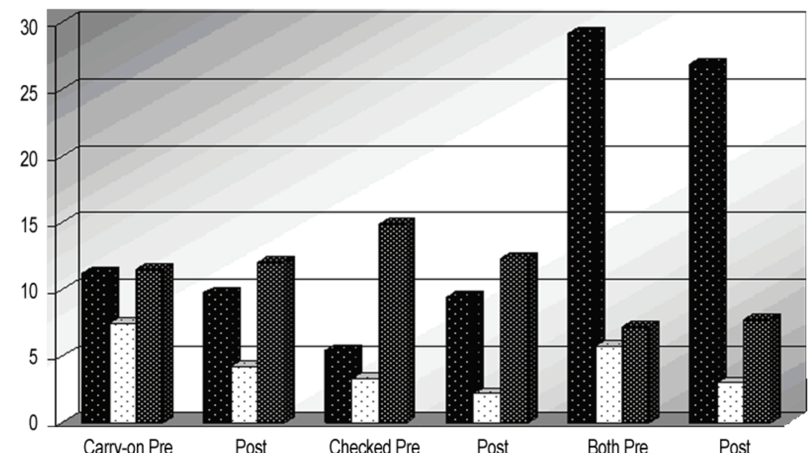
Impact of 9/11 on Air Travel

Since the events of 9/11 are what prompted this survey, several sections were devoted to determining its impact on travelers' behavior and attitude toward air transportation. The first section asked participants to rate the level of fear they had of flying within the US. The second section explored the use of carry-on versus checked baggage on business and personal travel before and after 9/11. The last section focusing on this area, highlighted the number of trips that were canceled after 9/11 through December 31, 2001. This included both business and pleasure air travel. The researchers what to compare the behavior of the respondents to their attitude about air transportation.

The 10-point Likert scale indicated 0 as No Fear and 10 as Terrified. Prior to 9/11 .5% of those who responded, were terrified to fly. After 9/11, that number increased to 2.5% or 19 out of the 764 respondents. However, prior to 9/11, 24.9% of respondents had no fear when flying within the US. Over 80% of the respondent rated their level of concern at 4 or lower. After 9/11, the ranking dispersed more evenly over the 10 point scale with only 49.7% giving their level of concern at a 4 or lower. Those with no concern decreased to 9.1% a drop of 15.8%. With this as a reference, a concern for airline safety caused 13.5% of respondents to not book a flight over the previous 12 months. Out of 766 respondents, 103 chose not to fly based on the perception of problems with airline safety. However, 662 chose to fly and one abstained from answering the question. Since the respondents were established to be frequent travelers, their response to 9/11 may not have been as severe as individuals more fearful or not as experienced with air travel.

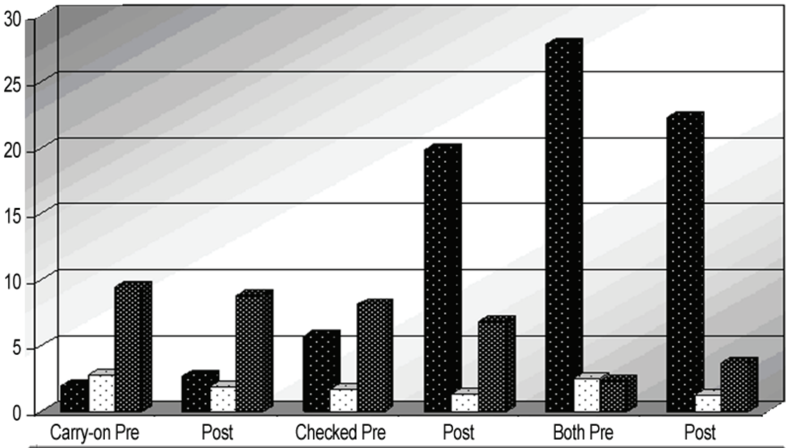
Carry-on and checked baggage became more of an issue after 9/11. Changes in airport screening and airline security have influenced the packing and behavior of travelers to ease their movement through the sometimes lengthy and intrusive process. Unfortunately, problems with checked baggage has made this a difficult process. In the last 12 months, 41.4% of the 702 that responded had their checked baggage lost or mishandled temporarily. However, 1.7% had their baggage lost and never returned.

The survey analyzed changes in traveler's preference regarding baggage check-in or carry-on usage following 9/11 in comparison to prior to the event. The preferences were further divided between business and personal travel (Figures 3 and 4). There has been little change in baggage preferences following 9/11. The respondents were given the option of checking all baggage versus carry-on of all baggage or a combination of the two. For the purpose of compressing data, the three main areas examined were when respondents indicated either 0%, 50% or 100%. In all categories, these were the primary responses selected. All other responses evenly covered the rest of the continuum. The most significant change occurred in the number of respondents on personal travel choosing to check in all of the baggage. The number rose from 11.4% to 39.6%. Travelers checking no luggage in this category dropped from 16.2% to 13.7%. Business travelers checking no luggage also dropped from 14.9% to 12.3%. However, business travelers checking all of their luggage also increased modestly from 5.4% to 9.4%.



	Carry-on Pre	Post	Checked Pre	Post	Both Pre	Post
100	11.2	9.7	5.4	9.4	29.1	26.8
50	7.4	4.2	3.3	2.2	5.7	3
0	11.5	12	14.9	12.3	7.2	7.7

Figure 2. Business Travel Baggage Preferences Pre-and Post-9/11



	Carry-on Pre	Post	Checked Pre	Post	Both Pre	Post
100	3.9	5.4	11.4	39.6	55.6	44.6
50	5.5	3.7	3.5	2.7	5	2.5
0	18.7	17.6	16.2	13.7	4.6	7.4

Figure 3. Personal Travel Baggage Preferences Pre-and Post-9/11

Business and personal travelers appear to have similar baggage preferences. Regardless of the travel purpose, the respondents indicated they preferred to use a combination of carry-on and checked baggage. This preference did not change with the new security procedures and safety programs established since 9/11.

The survey established little change in consumer baggage handling preferences. The final section dealing with 9/11 established any change in the respondents' travel schedule based on flights not taken or canceled due to the events. Following 9/11, 17.6%, 135 of the respondents, indicated that they canceled business travel plans and 12.9%, 99 of the respondents indicated they canceled pleasure travel plans. However, the overall majority did not cancel their trips—78.6% and 85.8% respectively. The questions further probed the number of trips canceled through December 31, 2001 and if the trips were accomplished using other travel modes.

Table 4
Impact of 9/11 on Canceled Travel and Unaccomplished Trips

	Type of Trip	
	Business	**Pleasure**
Number of trips canceled through 12/01	322	136
Percentage not accomplished at all	56.3%	71.7%

The figures from this portion of the survey are interesting in terms of trips canceled and never taken. However, there are limitations in the sense that the survey did not question as to whether the canceled travel occurred during the period when the air transportation was shut down or during its gradual recovery when many flights were cancelled even though there were willing travelers. In addition, travel could have been canceled due to other factors that were not completely reliant on 9/11. The numbers that are more remarkable are those detailing accomplished trips and travel continued even though the nation and air transportation was reeling from the images of 9/11.

Discussion

These findings provide a continued basis of validation for the AQR. While the survey represents the perceptions of frequent travel consumers, the AQR is based on quantitatively measured service indicators (Table 5). The relationship, while not a direct correlation, does show that there are similarities between these two different measures of overall quality of service. In three out of four of the AQR categories, the Airline Survey data indicated a comparable level of importance. The issue of involuntary denied boardings is the only area where there was a discrepancy between the two measures.

Table 5
Consumer Ranked Importance of Quality Matrix Links Airline Survey and AQR.

	Survey	**AQR**
On Time Arrivals	High	High
Involuntary Denied Boardings	Low	High
Mishandled Baggage	High	High
Combination of Customer Complaints	High	High
Impact of 9/11 on Air Travel	High	*

Note: * indicates not tested for in study

Periodically over the last decade, survey data such as those presented here document that there is a high degree of relationship between qualitative and quantitative measures of service quality. The fact that this relationship has continued over time and across multiple instruments provides a conclusive and convincing case of overall reliability to both types of measurement.

It is anticipated that these results will be cited in future AQR reports as further validation that the AQR measurement remains representative of consumer perceptions of quality. These perceptions, which have changed over time, have influenced the weighting applied to the AQR. However, these data make it clear that the AQR itself remains consistent with consumer expectations.

An interesting conclusion or final discussion point of this study is that the Airline Survey and AQR actually validate each other. This development was not part of the survey process. However, the data collected allowed the researchers to see similarities between the two apparatus. These similarities were found when the subjective determinants of service quality, revealed in the survey, matched the results of the AQR, which relies on quantitative performance driven measures. The performance drive data of the AQR also includes a subjective portion through its compilation of consumer-filed complaints. This data analysis serves as confirmation of the similarities between the two measures.

Conclusion

The Airline Survey of frequent travelers examined four distinctive areas of airline travel. Each one contributes insight into how travelers both perceive quality and select airlines. The events of 9/11 prompted this survey and their influence is assessed in several sections.

Little has changed in the perception of quality. The four major categories the AQR determined important to the traveler in 1991 still hold influence today. However, the safety records of the airline as well as competitive ticket pricing have gained in prominence. This may be due in part to the changes in society since the terrorist attacks of September 11, 2001 as well as economic issues. On-time arrivals, mishandled baggage, and management of customer complaint categories, maintain a prominent role according to the respondents' rating. Involuntary denied boardings does not seem to be as high of a concern. This may be due to changes to the airline booking strategies or the traveling habits of the public. Whatever the case, problems with boarding rate low in terms of importance to the respondents.

The second area provided insight into what factors go into selecting a specific airline. This included the level of importance participants assigned to key issues. For instance, having baggage arrive with the traveler ranked first with the safety record of airline a distant second. At the other end of the spectrum, the size and type of aircraft as well as the availability of upgradeable seats ranked low in importance.

The complaint category, with its range of options, remained significant. Given the level of importance respondents assigned to the handling of customer complaints and their actual rate of satisfaction, the Airline Survey's results suggested the travelers' have a low perception of quality airline performance. Complaints highlight issues of cancellation handling and delays. This closely relates to problems of rude treatment by airline employees. Lost luggage remains a significant problem.

The final area explored the impact on airline travel from the terrorist attacks of 9/11. The respondents indicated a 2% increase in the number of those terrified of air travel. However, 80% still have no fear.

The majority of air travel continued. Even with the slight change in baggage procedures, it is difficult to determine if this change is due to changes in airline baggage handling or in consumer preferences. With greater limitations and restrictions, the traveler has fewer options with his or her baggage than before 9/11.

Acknowledgements

This research was funded in part by the NASA Nebraska Space Grant Consortium and Nebraska NASA EPSCoR Program.

References

2005 AQR Airline Quality Rating (2005). *Airline Quality Rating*. Retrieved October 3, 2005, from http://www.aero.aqr

Airfares. Price it is. (2000, July). *Consumer Reports Travel Letter 16*, 3.

Berry, T., Bhadra, D. Gentry, J. & Nelsoni, G. (2004). *Airlines entry and exit and its impact on air traffic management: An analytical framework for zero-sum and positive sum games*. In S. Tarry (Ed.), *Transportation Research Forum. 45th Annual Research Forum Proceedings*. March 21-23, 2004. Evanston, IL.

Bisignani, G. (2002). *Airports and airlines: A need for a new partnership*. Speech to Airports Council International (ACI) by Giovanni Bisignani, IATA Director General & CEO. 8 October 2002. Retrieved on February 5, 2005 at http://www.iata.org/pressroom/speeches/2002-10-08-05.htm

Bisignani, G. (2004). *State of the air transport industry report by Director General*. Speech to International Air Transport of Association. June 7, 2004

Bitner, M. J. (1990, April). Evaluating service encounters: The effects of physical surroundings and employee responses. *Journal of Marketing*, *54*(2), 69-82.

Bowen, B. D. & Headley, D.E. (2004). *Airline quality rating 2004*. Retrieved on September 13, 2004 at http://www.unomaha.edu/~unoai/aqr/2004

Bowen, B, Headley, D., & Lu, C. (2003). Measuring airline productivity in the post 9-11 operational environment. *Proceedings in Air Transport Research Society annual meeting.* July 10-12, Toulouse, France

Chui, A., & Bowen, B. (2004). *Meta-analytic assessment of aviation forecasting.* Presented at the 104th annual meeting of the Nebraska Academy of Sciences. Lincoln, NE. April 13, 2004.

Cobb, R.W. & Primo, D.M. (2003). *The plane truth: Airline crashes, the media, and transportation policy*. Washington, DC: Brookings Institution Press.

Embraer Commercial Jets. (2004). *The rule of 70 to 110. Strategy for right-sizing from Embraer Commercial Jets*. Retrieved on September 13, 2004 from http://www.Rulef 70to110.com.

Fick, G. R. & Ritchie, B. (1991, Fall). Measuring service quality in the travel and tourism industry. *Journal of Travel Research, 30*(2), 2-9.

Fink, A. (1995). *The survey handbook: The survey kit book 1*. Thousand Oaks, CA: Sage.

Frequent Flyer (1997). The people's choice. *Frequent Flyer*, June, pp. 24-26.

Ghobrial, A. & Irvin, W.A. (2004). Combating air terrorism: Some implications to the aviation industry. *Journal of Air Transportation 9*(3), 67-86.

Headley, D. & Bowen, B. (1997). International airline quality measurement. *Journal of Air Transportation World Wide 2*(1), 55-63.

Kane, R.M. (2003). *Air transportation, 14th edition.* Dubuque, IA: Kendall/Hunt Publishing Company.

Office of Aviation Enforcement and Proceedings (2005) *Air travel consumer report.* U.S. Department of Transportation: Washington, DC. Retrieved on April 2, 2005 from http://airconsumer.ost.dot.gov/reports/

Office of Aviation Policy & Plans. (2003, March). *FAA aerospace forecasts: Fiscal years 2003-2004*. U.S. Department of Transportation, Federal Aviation Administration [FAA], & Office of Aviation Policy & Plans: Washington, DC.

Oster Jr., C.V., Strong, J.S., & Zorn, C.K. (2000/2001). Improving air safety: Long-term challenges. *Issues in Science & Technology 17*(2), 77-83.

Ott, J. (1998, November 9). Technology pressures build for new slant on aviation agreements. *Aviation Week & Space Technology, 65*, 68-69

Rhoades, D.L. & Waguespack, Jr., B. (2000). Service quality in the U.S. airline industry: Variations in performance within airlines and between airlines and the industry, *Journal of Air Transportation World Wide* (5), 60-77.

Rupp, N.G., Holmes, G.M., & DeSimone, J. (2003). *Airline schedule recovery after airport closures: Empirical evidence since September 11th.* National Bureau of Economic Research Working Paper No. 9744. Issued June 2003. Retrieved on February 8, 2005 at http://papers.nber.org/papers/w9744

Scarpellini-Metz, N., & Bowen, B. (2004, October). *Identifying synergistic relationships of National Aviation (Blue Ribbon) Commission Reports: A qualitative data analysis application.* Presented by Bowen at University Aviation Association Fall Meeting, October 6-9, 2004. Toronto, Canada. 20 pps.

Stephenson, F.J. & Bender, A.R. (1995, Spring). Watershed: The future of US business travel. *Transportation Journal 35*(3), 14-32.

The Boyd Group. (2004a, October 9). *Forecast airline industry trends strategies.* Presentation at The Boyd Group 9th Annual Aviation Forecast Conference, Denver, CO.

The Boyd Group. (2004b, October 10). *Forecast fleet trends.* Presentation at The Boyd Group 9th Annual Aviation Forecast Conference, Denver, CO.

The Boyd Group. (2004c, October 10). *Passenger forecast 2004-2010*. Presentation at The Boyd Group 9th Annual Aviation Forecast Conference, Denver, CO.

International Journal of Applied Aviation Studies, 5, Number 2

Utility of a Personal Computer-Based Aviation Training Device for Helicopter Flight Training

David M. Johnson and John E. Stewart II
U.S. Army Research Institute for the Behavioral and Social Sciences, Rotary Wing Aviation Research Unit, (DAPE-ARI-IR), Building 5100, Fort Rucker, AL 36362-5354

david.johnson@rucker.army.mil, 334.255.3915
john.stewart@rucker.army.mil, 334.255.9109

Abstract

Personal computer-based aviation training devices (PCATDs) have been shown to support primary flight training. Positive results have been shown for fixed-wing aircraft only. This research investigated which tasks from U.S. Army Initial Entry Rotary Wing (IERW) training could be supported by a PCATD. Sixteen aviators representing both highly experienced and student helicopter pilots evaluated the ability of a commercial PCATD, running Microsoft Flight Simulator 2000™ Professional Edition, to support the seventy-one flight tasks comprising the IERW Common Core curriculum. Aviators performed each task one or more times in the PCATD before rating it on a four-point scale. Additional data were gathered on attitudes toward simulation and computer literacy, as well as comments and criticisms. Results showed remarkable agreement between experienced aviators and students. The PCATD was judged as best supporting Instrument Flight Training, especially tasks involving radio navigation. Tasks from Primary Flight Training, especially tasks requiring hovering, were judged as less well supported. The most frequently stated positive comment was that the PCATD would be valuable in training navigation instruments and procedures. The three most frequent criticisms concerned narrow field of view, poor visual depth cues, and inability to perform hovering flight tasks.

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

Introduction

According to a number of sources, consumer flight sims [simulators] have become a de facto part of Air Force flight training. "It's almost like that's the first phase of training—you come here fully trained up on [Microsoft] flight simulator and we'll throw you into an Air Force simulator and see how you handle it" (Prensky, 2001, p. 310).

What is a PCATD?

It is widely recognized that both the power of PC processors and the capability of consumer flight simulation software have grown dramatically at the same time that they have become less expensive. Both the private and the military aviation communities have noticed this trend and are seeking ways to harness it in service to their own training needs. This has lead to the personal computer-based aviation training device (PCATD). In their review of this technology, Koonce and Bramble (1998) noted that PCATDs run the gamut from desktop devices to flight simulators complete with cockpit shells and full instrumentation.

In the aviation training literature, the acronym PCATD usually refers to a relatively inexpensive, fixed-platform flight simulator based on microchip processing technology and commercial-off-the-shelf, consumer-oriented software. These devices run on standard desktop PC-based operating systems such as the various versions of Microsoft Windows™. The out-the-window view is shown on the host computer's display screen. Usually, flight and navigation instruments are also shown on the host's screen as well. Some systems provide separate display screens for instruments or actual instruments mounted in a simulated cockpit shell. Although the consumer software allows flight to be controlled from a standard PC keyboard, this is seldom used in the training literature. Flight controls are usually separate hardware consisting of some combination of flight stick and rudder pedals, control yoke and rudder pedals, cyclic and pedals (for helicopter applications), as well as a throttle or collective with throttle (helicopter). Some systems are sold as a complete cockpit with a seat, instruments, controls, large screen, and separate speakers; while others rely on the user's chair, the host PC's built-in speaker, screen, and separate add-on flight controls. Most PCATDs sell for less than $10,000--often much less--making them orders of magnitude less expensive than the flight simulators historically sold by major manufacturers for civilian or military training applications.

Research Involving PCATDs

Research, both in the private sector and the military, has investigated the value of using PCATDs for fixed-wing flight training. Ortiz, Kopp, and Willenbucher (1995) investigated the effectiveness of PCATDs for training instrument flight skills in a sample of 26 students at the Lufthansa Pilot's School. Two matched groups of students were compared. One group received part of its instrument training in a PCATD while the other group received the standard course of instruction using an approved procedures trainer. No statistically significant differences in flight performance were observed between the two groups at the end of course check ride. The total cost of the PCATD, however, was only three percent of the cost of the certified procedures trainer.

Taylor, Lintern, Hulin, Talleur, Emanuel, and Phillips (1997) in a large, well-designed study reported positive transfer of training from PCATDs to the aircraft for instrument flight tasks. Their experiment involved 144 participants taking instrument flight training courses at the University of Illinois. They measured both trials to criterion in the aircraft and time to criterion in the aircraft. Using the data the authors calculated percent transfer and transfer effectiveness ratios (TERs). They found statistically significant positive transfer in seven flight lessons with percent transfer ranging from 15% to 33% and TERs ranging from 0.16 to 0.39. These results imply that for every 2.5 to 6 trials of training in the PCATD, 1 trial in the Beechcraft Sundowner was saved. Taylor et al. concluded that the PCATD is an effective training device for teaching instrument tasks to private pilots.

Dennis and Harris (1998) reported positive transfer of training from desktop PCATD practice to the aircraft for a small sample of simple visual flight maneuvers. These authors employed 21 participants, none of whom had had any prior flight experience. The members of the PCATD group received a total of one hour of training on the desktop simulator prior to their test flight in a Beagle Pup 121 airplane. Compared to a control group that did not receive the simulator training, the PCATD group demonstrated significantly better performance in straight-and-level flight as well as in exiting turns. There are two reasons for mentioning this experiment. First, transfer was shown for visual flight tasks. Second, this significant positive transfer was shown after a single hour of training with the PCATD.

Koonce and Bramble (1998) reviewed the history and present status of the use of PCATDs for fixed-wing flight training. They made a number of separate points that will be paraphrased here. First, instrument training has long been an area of flight training that has benefited from simulation. Readers are no doubt aware that the Link Trainer or “blue box” began institutional use prior to World War II. Second, Koonce and Bramble reviewed Air Force research that showed as early as 1966 that procedural tasks, such as instrument flight tasks, were far more likely to transfer well to the airplane than were perceptual-motor flight tasks, such as stick and rudder skills. Third, simulator technology continues to improve. Thus, what appear to be “aviation games” today (i.e., PCATDs), surpass in power and fidelity what were state-of-the-art full-mission simulators in 1970. Finally, the authors reported two studies that showed positive transfer of training from PCATDs to airplanes for instrument tasks and one study that showed positive transfer from PCATDs to airplanes for visual flight rules (VFR) tasks. The point is that PCATDs appear to be following in the footsteps of earlier flight simulator technology. They begin as games, and then demonstrate their validity for instrument flight training, and finally move on to simulating visual (contact) flight for ground-school institutional training.

In keeping with this trend, and based in part on the Taylor et al. (1997) research, the Federal Aviation Administration (FAA) authorized the use of PCATDs to satisfy part of the creditable training requirement for the private pilot instrument rating. The details of this approval were described in FAA Advisory Circular #61-126 (1997). The key provisions were these: If the PCATD meets FAA guidelines, and if the instruction is by an FAA Certified Flight Instructor, then PCATD instruction can be used to satisfy as many as 10 of the 20 simulator hours creditable for instrument rating in part 61 (or as many as 10 of the 15 simulator hours creditable for instrument rating in part 141) of FAA Regulations.

In order for a PCATD to qualify for FAA approval, extensive physical controls must be present. These include: A physical, self-centering displacement yoke or control stick that allows continuous adjustment of pitch and bank; physical, self-centering rudder pedals that allow continuous adjustment of yaw; a physical throttle lever or power lever that allows continuous movement from idle to full power settings; and physical controls for as many as 12 other items applicable to the aircraft being replicated, such as flaps and navigation radios for example. Interestingly, Koonce and Bramble (1998) stated that they could find no scientific evidence supporting the requirement for the 12 additional physical controls mentioned above.

The Chief of Naval Education and Training began a project to identify and apply commercial PC gaming and simulation technology as a potential training tool. Dunlap and Tarr (1999) reported one outcome of this project that generated considerable interest in military training circles. They configured 10 simulator workstations as Navy T-34C fixed-wing training aircraft. Fifteen scenarios were developed including familiarization flights, basic instruments, and navigation instruments. Microsoft Flight Simulator 98™ software was augmented with an instructional framework that provided a demonstration of each scenario narrated to point out key visual and timing events. After the demonstration, student pilots were afforded the opportunity to practice the scenario. All participation in this training experience was voluntary and performed at a time that would not interfere with the primary flight curriculum. Results from this initial test were positive. Participating students were significantly more likely to score highly during flight training and significantly less likely to “wash out” of flight training, when compared to their peers who did not participate. However, this initial study did not contain a control group—a fact freely admitted by the authors. Thus, it is impossible to know if practice with the PCATD helped the participating students, or if the best students were the ones who chose to participate.

Schneider, Greene, Levi, and Jeffery (2001) of the Air Force performed a controlled experiment comparing standard flight instruction to standard flight instruction plus PCATD practice. They compared the flight training performance of 55 students who were provided with access to PCATDs with that of 209 students who received standard training. The two groups of students were compared on their learning of nine advanced contact flight maneuvers such as loop and barrel roll. There were two measures of trials to criterion for each task and two measures of variability for each task. Although performance on all four measures favored the PCATD group in absolute magnitude, the small differences were statistically significant in one case only. Overall, the PCATD group showed significantly less variability than the control group.

Roessingh (2005) investigated the effects of pretraining in two simple PC and PCATD devices on the performance of five standard aerobatic maneuvers in a light aircraft (Bellanca Super Decathlon). Trainees comprised three groups: a control group that received no pretraining, and two experimental groups. Of the latter, one trained on a PC equipped with simple spring-loaded joystick controls, standard furniture and a 43 cm color monitor; the second group trained on a PCATD equipped with spring-loaded controls and seat physically and functionally similar to those of the target aircraft, and adjustable 53 cm color monitor. Both PC and PCATD conditions employed a commercially available software package (Looking Glass Technologies, 1995) that allowed aerobatic maneuvers to be practiced on a desktop PC. Trainees in the PCATD condition had additional instructional options in the software package that provided corrective feedback for each maneuver. In the aircraft, maneuvers were recorded via in-flight data recording equipment. These data were analyzed post flight using predetermined, objective criteria. Comparison of the three groups revealed no significant differences in aerobatic performance. However, experimental participants seemed to show advantages in procedural training, requiring less preflight briefing time than controls. Whereas control participants required approximately 15 min of briefing time prior to flight, PC participants required approximately 5 min, and their PCATD counterparts required virtually no briefing time. In addition, experimental participants seemed to use their training time more efficiently than controls, performing significantly more maneuvers per lesson. Thus, it seems that these simple PC-based trainers showed more promise for training procedural skills than for the perceptual-motor skills required for executing complex flight maneuvers. Although Roessingh stated that overall there were no significant differences between groups in terms of the execution of the trained maneuvers, it nevertheless appeared that the experimental groups were better prepared for their lessons because of pretraining.

In summary, the fixed-wing research reported above appears to support two conclusions. First, PCATDs have been shown to be an effective medium for training the procedural skills required for instrument flight (Dunlap & Tarr, 1999; FAA, 1997; Koonce & Bramble, 1998; Ortiz et al., 1995; Taylor et al., 1997). Second, PCATDs have not shown promise for training of the perceptual-motor skills required for aerobatic visual flight maneuvers (Roessingh, 2005; Schneider et al., 2001).

Recent research has investigated the use of PCATDs for tasks other than the basic flight training of novices. Talleur, Taylor, Emanuel, Rantanen, and Bradshaw (2003) examined the effectiveness of PCATDs for the maintenance of instrument currency among 106 instrument-rated pilots. There were three experimental groups and one control group. All groups received an initial instrument proficiency check (IPC) followed 6 months later by a second IPC. Both IPCs were administered in a Beechcraft Sundowner airplane. The three treatment groups received two instrument training sessions between IPCs. The training sessions took place in an FAA-approved PCATD (Jeppesen FS-200), or in an FAA-approved flight training device (FTD, Frasca 141), or in the airplane. The control group received no instrument training between IPCs.

The results showed that while there were no statistically significant differences among the groups at IPC1, there were at IPC2. A measure of the IPC pass/fail rates showed that both the PCATD group and the FTD group performed significantly better than the control group. There was no significant difference between the airplane group and the control group. The authors concluded that PCATDs are effective for maintenance of instrument currency. They also reported, "PCATD training was beneficial for all critical instrument maneuvers that involved procedural components, such as VOR navigation, ILS approaches, and holding procedures." (Talleur et al., p. 398).

In related research Taylor and colleagues (Taylor, Talleur, Emanuel, & Rantanen, 2004; Taylor, Talleur, Rantanen, & Emanuel, 2005) investigated the effectiveness of PCATDs for conducting IPCs. The purpose of this experiment was to compare the performance of pilots who received an initial IPC in a PCATD, an FTD, or an airplane. There were three groups of 25 pilots each. Pilots were randomly assigned to one of three initial IPC conditions: PCATD (FAA-approved Elite), FTD (FAA-approved Frasca 141), or a Beechcraft Sundowner airplane. All groups received a second IPC 2 weeks later in the airplane.

The results showed no statistically significant differences among the groups either at IPC1 or at IPC2. There was no evidence at IPC2 that performance was improved by the use of the more expensive FTD or airplane for the initial IPC. The authors argued that these results supported the option of using PCATDs for IPCs.

PCATDs have also been used in research that is not directly related to flight training. They have been employed as platforms for research in aircrew coordination, crew resource management (CRM), and collective performance of aircrew teams. Jentsch and Bowers (1998) reviewed the early work in this area. A recent article describing a quasi-transfer of training experiment employing PCATDs to investigate CRM issues among helicopter crews was reported by Brannick, Prince, and Salas (2005). Also recently, Proctor, Panko, and Donovan (2004) used networked PCATDs to examine the performance of rated Army aviators in multi-helicopter operations.

Purpose of this Investigation

The flight-training research described above was all performed using PCATDs to simulate fixed-wing aircraft. To date the authors have uncovered no published research employing PCATDs in introductory helicopter flight training. This explains the need for the current study.

Desk Top Simulators L.L.C. of Fort Worth, Texas loaned two identical devices, called Rapidly Transferable Cockpits (RTCs), to the Directorate of Training, Doctrine, and Simulation (DOTDS) at Fort Rucker for an initial 90-day examination period. The Army Research Institute for the Behavioral and Social Sciences (ARI) was asked to evaluate these two PCATDs. Throughout this report, the devices will be called "micro-simulators" in keeping with the usage established by Dunlap and Tarr.

Bell and Waag (1998) listed three categories of approach for evaluating the training effectiveness of flight simulators. Utility evaluations are the easiest and quickest. Subject matter experts perform specific tasks or missions in the simulator and then rate the effectiveness of the simulator for training. The second category is in-simulator learning. Novices practice tasks in the simulator and thereby show learning through an improvement in performance. Typically the method is one of pre-test, practice, and then post-test. In this case, practice in the simulator can be shown to produce an improvement in performance in comparison to an appropriate control group. The third category is transfer of training. Here the trainee is transferred to a new environment, such as an actual aircraft, after training in the simulator. The goal is to show that the skills learned in the simulator improve performance in the aircraft in comparison to a control group not pre-trained in the simulator. Transfer of training is an excellent method to evaluate the training effectiveness of a simulator although it is resource intensive—requiring students, instructors, aircraft, and time.

The present investigation was a utility evaluation, a rapid test of an existing micro-simulator by users (cf., Nielsen, 1993). ARI was primarily interested in discovering what Initial Entry Rotary Wing (IERW) tasks could be supported for training by this device. This investigation was part of a larger examination of flight training at the U.S. Army Aviation Center (USAAVNC) at Fort Rucker. ARI chose to employ the technique of a utility evaluation in order to provide USAAVNC with reliable information in a timely manner.

The evaluation proceeded in two phases that were identical in all respects except for the category of aviator doing the behavioral evaluation. Phase 1 included experienced aviators, phase two, student aviators. This allowed the micro-simulator to be evaluated by people who were representative of the instructors and the students who would potentially use it. In brief, an attempt was made to have the micro-simulator under investigation evaluated by members of the target audience of interest to USAAVNC. This issue of the target audience is important. The history of human factors engineering is rife with examples of very powerful systems that saw limited use because they were designed without input from the end user and were, therefore, inappropriate for the task or incapable of being used by the intended audience (Nielsen, 1993; Norman, 1988; 1998). ARI was determined not to make this mistake.

A decision to employ 16 evaluators (6 experienced + 10 students) was based in part on the advice of Nielsen (1993). Research by Nielsen concerning the optimal number of participants for usability testing has shown that 15 are sufficient. Ninety percent of the usability problems to be found will be found by 15 evaluators. Sample sizes larger than this increase testing costs without producing significant increases in benefits.

Method

Phase One Participants: Experienced Aviators

Six experienced helicopter pilots volunteered to evaluate the micro-simulator. All were male. (Demographically, the vast majority of Army aviators are males.) They ranged in age from 33 to 55 years, with a mean age of 43.2 years (median 43 years). All were rated Army pilots. Their total aircraft flight hours ranged from 1153 to 5500 with a mean of 2742.2 hours (median 2400). Five were current or former instructor pilots (IPs), current or former standardization IPs, or current or former maintenance test pilots. All were rated in various models of the UH-1, 5 in various models of the OH-58, 4 in various models of the AH-1, 4 in the TH-55, 2 in various models of the AH-64, and 1 each in the OH-13, the TH-67, and various single-engine, piston-powered Cessna aircraft. At the time of the evaluation, the U.S. Army at Fort Rucker employed all six in some training-related capacity.

Phase Two Participants: Student Aviators

Ten student pilots volunteered to participate in this phase of the evaluation. Nine were male and one was female. Four were commissioned officers (2LT) and six were warrant officers (WO1). They ranged in age from 20 to 31 years with a mean age of 26.4 years (median 27.5 years). Their total aircraft flight hours ranged from 78 to 500 with a mean of 144.6 (median 120). All 10 had completed Primary Flight Training Stage I, Primary Flight Training Stage II, Instrument Flight Training Stage I (Basic Instruments), and Instrument Flight Training Stage II (Advanced Instruments). Five were awaiting assignment to the last phase of IERW training called Basic Combat Skills. The remaining 5 had completed Basic Combat Skills and were awaiting assignment either to the Officer's or the Warrant Officer's Basic Course. Thus, all student evaluators had just completed that portion of the flight-training curriculum for which the micro-simulator in question was being evaluated for possible use.

Simulator

Each RTC weighed approximately 227 kg. Its dimensions were 183 cm in length, 86 cm in width, and 152 cm in height. It used standard 110 V, 60 Hz power. The visual display monitor measured 71 cm diagonally. The angular field of view of this screen from a normal sitting position was 43° (horizontal) by 34° (vertical). This CRT screen had a resolution of 768 pixels horizontally by 1024 lines vertically. The RTC was capable of supporting a wide variety of PC-based flight simulator software applications. This evaluation was limited to Microsoft Flight Simulator 2000™ Professional Edition employing the Bell 206B Jet Ranger. The software was run on a Microsoft Windows 98™ operating system. The host computer was an Intel Pentium III™ Processor, operating at a speed of 550MHz, with 256MB RAM. The system included a Logitech wireless keyboard, a Logitech wireless mouse, and two Juster Multimedia speakers. The RTC contained a padded seat facing the CRT screen and operational flight controls. The cyclic was a Stick II made by Flight Link. Flight Link also made the collective and the pedals. Figures 1 and 2 show a lateral view and a rear view, respectively, of an RTC emplaced at the ARI test facility on Fort Rucker.

Figure 1. Lateral view of commercial micro-simulator PCATD.

Figure 2. Rear view of micro-simulator showing visual display.

IERW Flight Tasks

Seventy-one IERW flight tasks were chosen for evaluation. All tasks were taken from IERW Flight Training Guides (USAAVNC, November 1999; USAAVNC, February 2001). Twenty-nine tasks were selected from Stage I of Primary Flight Training. Six tasks were selected from Stage II of Primary Flight Training. Thirteen tasks were selected from Stage I (Basic Instruments) of Instrument Flight Training. Twenty-three tasks were selected from Stage II (Advanced Instruments) of Instrument Flight Training. These 71 total tasks constitute virtually the entire portion of the specifically flight-oriented curriculum. Non-flying tasks were not included.

Primary Flight Training is the student's first experience with flying the aircraft. Primary Flight Training tasks are VFR tasks. The student aviator is using his or her out-the-window view and ground references to control aircraft maneuvers. Examples of such tasks are "straight-and-level flight" and "takeoff to a hover."

In instrument flight rules (IFR) flight, the student aviator is primarily using his or her flight and navigation instruments as references to maintain control of aircraft attitude, altitude, and direction. Basic Instruments (BI) training concerns the use of flight instruments to control the aircraft and maintain proper attitude in the absence of visual cues. Advanced Instruments (AI) training concerns the use of radio navigation instruments in addition to, and in conjunction with, flight instruments. Examples of tasks taught as part of BI are standard rate turns and unusual attitude recovery. Examples of tasks requiring navigation instruments are Automatic Direction Finder (ADF) course tracking and Very High Frequency Omni-Directional Radio Range Receiver (VOR) approach.

Procedure

All participants were briefed on their role in this evaluation. They were asked to evaluate the micro-simulator in terms of how well it supported the 71 specific IERW Common Core flight tasks. They were asked to provide one evaluation for each task. The four possible levels of evaluation for each task were: "Not at all," "Slightly," "Moderately," and "Well."

Each participant was given an opportunity to operate ("fly") the micro-simulator until he or she was comfortable with it. When the aviator was ready, the formal evaluation period began. The tasks were listed on the questionnaire in the order they appeared in the Flight Training Guides (FTGs). Usually the ARI researcher named a task from the questionnaire, the participant performed it in the simulator one or more times, and then offered an evaluation score. FTGs were available if an aviator wanted to check the published performance criteria prior to attempting a task, and many did so. Many aviators, however, performed the tasks in their own idiosyncratic order calling out to the researcher what they were doing and then, when finished, offering their evaluation. Thus task presentation order was not held constant across all participants and so no order effects are implied by the results. Evaluators frequently verbalized their concerns or comments while operating the simulator. The researcher noted these concerns and comments on paper and asked questions when necessary. Although an effort was made to have all 16 aviators evaluate all 71-flight tasks, this was not always possible because of time constraints.

After the flight task evaluation portion each aviator was asked some general questions about simulation, PC flight simulators, personal PC usage, and how best to use this micro-simulator in IERW training. Two short breaks were taken during the evaluation period. The entire procedure required 4 hours of concentrated work.

Results

Flight Task Evaluation

Data were aggregated by task and by participant both for the experienced aviators and for the students. Mean and median ratings were calculated for each task. An evaluation of “Not at all” was rated 0, an evaluation of “Slightly” was rated 1, an evaluation of “Moderately” was rated 2, and an evaluation of “Well” was rated 3.

The mean overall evaluation given by the experienced aviators for all 71 tasks was 1.72. For student aviators the mean overall evaluation was 2.00. This small difference in overall mean ratings was statistically significant (nonparametric: Mann-Whitney U test, $N = 142$, $z = 2.84$, $p < .005$). The difference in overall median ratings was also statistically significant (nonparametric: Mann-Whitney U test, $N = 142$, $z = 2.34$, $p < .02$). Students consistently tended to rate the micro-simulator a little higher than did the experienced aviators in its ability to support IERW training.

Student ratings of the flight tasks were positively and significantly correlated with those ratings provided by the experienced aviators. This was true both for mean ratings (nonparametric: Spearman rank order correlation, $r_s = 0.78$, $N = 71$, $p < .001$) and for median ratings (nonparametric: Spearman rank order correlation, $r_s = 0.79$, $N = 71$, $p < .001$). That is, though the students tended to assign somewhat higher ratings, there was substantial agreement between the students and the more expert pilots as to which of the 71 flight tasks were well supported by the micro-simulator and which were not. These results are shown more clearly in Table 1.

Table 1 presents the results of the evaluations summarized across the four stages of flight training (Primary Stage I & II, Instrument Stage I & II) and the two categories of aviator (experienced & students). Both experienced and student aviators rated the micro-simulator as better able to support Instrument Flight Training than Primary Flight Training. Both experienced pilots and students rated the micro-simulator as best able to support Advanced Instruments. Overall, the micro-simulator was evaluated as “slightly” able to support Primary Flight Training but “moderately” or “well” able to support the Instrument Flight Training stages. This difference in ratings favoring the ability of the micro-simulator to support Instrument tasks over Primary tasks was statistically significant both for experienced aviators (nonparametric: Sign test, $N = 6$, $X = 0$, $p < .02$) and for students (nonparametric: Sign test, $N = 10$, $X = 0$, $p < .001$).

Table 1
Summary of Evaluation Ratings by Stage of Flight Task

Stage of Flight Training Tasks	Evaluators			
	Experienced		Students	
	Mean	Median	Mean	Median
Primary Flight Training Stage I	1.19	1	1.34	1
Primary Flight Training Stage II	1.29	1	0.91	1
Instrument Flight Training Stage I (Basic)	2.07	2	2.43	3
Instrument Flight Training Stage II (Advanced)	2.32	2	2.89	3

These results can be further explained by a listing of the specific tasks that were rated as best supported by the micro-simulator and those that were rated as least supported. Table 2 presents the top quartile of tasks (highest-rated 25%) as judged by both experienced and student evaluators. Table 3 presents the bottom quartile of tasks (lowest-rated 25%) as judged by both groups of aviators. No tasks were included in either table for which there were fewer than four experienced evaluations or six student evaluations.

Table 2
Top Quartile: Tasks Judged to be Best Supported by the Micro-Simulator

Task	Evaluators	
	Experienced	Students
Primary Flight Training Stage I		
Go-around	X	
Primary Flight Training Stage II		
Simulated maximum performance takeoff	X	
Instrument Flight Training Stage I (Basic)		
Straight-and-level flight		X
Straight-and-level flight (EP)	X	
Compass turns (EP)	X	X
Instrument Flight Training Stage II (Advanced)		
ADF station identification	X	X
ADF aircraft position	X	
ADF course interception	X	X
ADF course tracking	X	X
ADF holding	X	X

ADF approach	X	X
ADF missed approach	X	X
Radio navigation		X
VOR station identification	X	X
VOR aircraft position	X	X
VOR course interception		X
VOR course tracking		X
VOR approach		X
LOC/ILS station identification	X	X
LOC/ILS course interception	X	X
LOC/ILS course tracking	X	X
LOC/ILS approach	X	X
LOC/ILS missed approach	X	

Table 3
Bottom Quartile: Tasks Judged as Least Well Supported by the Micro-Simulator

Task	Evaluators	
	Experienced	Students
Primary Flight Training Stage I		X
Hover check	X	
Hover power checks	X	X
Hovering flight	X	X
Takeoff to a hover	X	X
Hovering turns	X	X
Landing from a hover	X	X
Rectangular course		X
Traffic pattern flight		X
Traffic pattern entry	X	
Traffic pattern exit	X	
VMC takeoff (hover)	X	
VMC (normal) approach (hover)	X	X
Approach termination procedure	X	X
VMC (normal) approach (ground)		X
Standard autorotation	X	X
Hovering autorotation	X	X
Simulated engine failure at hover altitude	X	X
Primary Flight Training Stage II		
VMC (steep) approach	X	X
Slope operations	X	X
Low-level autorotation	X	X
Standard autorotation with 180-degree turn	X	X

Of the 18 tasks rated by the experienced aviators as best supported by the micro-simulator, 16 were part of the Instrument Flight Training curriculum. Of these, 14 were trained as part of AI and involved radio navigation. For the student aviators the pattern of evaluations was much the same. All 18 tasks rated as best supported by the students were part of Instrument Flight Training. Of these, 16 were trained as part of AI and involved radio navigation. Thirteen tasks were rated highest by both groups of evaluators. Of these, 12 were a part of the radio navigation curriculum. Clearly, as previously shown in Table 1, aviators sampled in this evaluation rated the micro-simulator highest in its ability to support the AI stage of IERW.

The bottom quartile consists of the 18 tasks rated as least well supported by the micro-simulator. The list of these tasks is found in Table 3 for both experienced aviators and students. All the tasks listed in Table 3 were from Primary Flight Training. Fifteen tasks were rated as poorly supported by the micro-simulator by both groups of evaluators. Eleven of these were part of Stage I Primary Flight Training and the remaining four were from Stage II. As also shown in Table 1, aviators sampled in this evaluation rated the micro-simulator as *least* able to support Primary Flight Training. Note also that of the 21 tasks listed in Table 3, 15 involved hovering. (This includes the ten with the word "hover" in their title plus approach termination procedure, standard autorotation, slope operations, low-level autorotation, and standard autorotation with 180-degree turn. These tasks all have a hovering component.) Clearly, the aviators sampled in this evaluation gave low marks to the micro-simulator in its ability to support tasks requiring hovering.

General Questions

Four questions were asked of both groups of aviators after the flight task evaluation was completed. The first question was "All in all, I believe that simulation is an effective tool for initial flight training." Participants indicated their agreement with this statement by checking one of six boxes along a (6-point) Likert-type scale. The choices were "Strongly Agree," "Agree," "Agree Somewhat," "Disagree Somewhat," "Disagree," and "Strongly Disagree." For both experienced and student aviators the mean and median response was "Agree." In other words, both groups of aviators agreed that, in principle, simulation is an effective tool for initial flight training.

The second question asked whether the participants had run any desktop flight simulator or aviation-related game on a PC in the past year. All six experts said "yes." Only 4 of the 10 students said "yes."

The third question assessed whether or not the participants had access to a PC at their place of residence. All six experts answered "yes." Nine of the 10 students answered "yes."

The last question was an open-ended one. Participants were asked, "If you were in a decision-making capacity, how would you employ this micro-simulator for IERW flight training?" These answers were entirely in keeping with their flight task ratings. Fifteen of the 16 evaluators expressed the opinion that the micro-simulator had value for Instrument Flight Training. Interestingly, four of the experienced and two of the students used the term "procedures," "procedures trainer," or "instrument techniques" in answering the question. Clearly, the respondents were attempting to communicate to the researcher that the machine had value as a way of practicing specific procedures for heavily procedures-driven instrument tasks.

One experienced aviator and one student expressed the opinion that the micro-simulator would be of value as a dynamic classroom teaching aid in ground school. Three students stated that the machine would be of value as a practice tool and should be available to flight students in the Learning Center at Fort Rucker.

Comments and Criticisms

Aviators spontaneously reported their comments and criticisms of the micro-simulator while performing the flight task evaluations. Comments and criticisms from all 16 evaluators were aggregated by common themes. A comment or criticism was noted only if stated by at least two evaluators. Evaluators reported a total of 5 positive comments and 14 criticisms.

The most frequently stated positive comment was that the micro-simulator was an instruments trainer. It would be valuable in supporting the training of navigation instruments and procedures during Instrument Flight Training. Fourteen evaluators mentioned this theme.

The three most frequently cited criticisms of the micro-simulator concerned visual field of view (FOV), visual cues to depth, and inability to hover. All 16 evaluators commented that a helicopter flight simulator must have a wider FOV than was provided. Peripheral visual cues are required for hovering tasks, traffic pattern flight, traffic pattern entry/exit, autorotation, and other VFR tasks. Fourteen evaluators stated that the visual cues to height above ground must be improved for VFR tasks. Helicopter pilots use out-the-window visual cues to determine height above terrain for a wide range of VFR tasks such as hovering, approach, and autorotation. None of the aviators could achieve a stable hover with this micro-simulator. Twelve evaluators commented upon this deficiency directly. The simulator was reported to be much harder to hover than the helicopter itself.

Discussion

This was an evaluation of the utility of a commercial micro-simulator running a PC-based flight simulator application (Microsoft Flight Simulator 2000™). It was a behavioral evaluation in the sense that aviators performed each flight task listed at least once before providing a rating. The sample of aviators who rated the device was representative of the target audience of instructors and students of value to USAAVNC. This was *not* a training experiment. No novice flight students were trained to criterion using this micro-simulator and then compared to some relevant control group. Further, it would be technically incorrect to generalize these findings to another micro-simulator running a different software application. (For example, Microsoft Flight Simulator 2004™ is now available, and would have to be evaluated separately.) However, within the constraints imposed by these limitations certain clear conclusions can be drawn.

First, there was remarkable agreement between the experienced aviators and the students in their flight task evaluations, their answers to the general questions, their positive comments, and their criticisms. This was shown by the statistically significant correlations reported for the ratings. This agreement between high-time aviators and student aviators has practical consequences. It suggests that the results of this particular evaluation were so obvious that one did not need to be an expert to notice them. In addition, it suggests that sometimes relatively easy-to-get student aviators might be substituted for relatively difficult-to-get subject matter experts in simulator evaluation research.

Second, the ratings argued persuasively that this micro-simulator could be used to support Instrument Flight Training in some capacity. Evaluators stated that both instrument flight tasks and navigation tasks could be trained to some extent using it. Those procedural, knowledge-based, radio navigation tasks that are trained as a part of AI, in particular, were judged as well supported. This conclusion drawn from Army rotary-wing aviators was consistent with prior fixed-wing research by the Navy (Dunlap & Tarr, 1999), by the private sector (Koonce & Bramble, 1998; Ortiz et al., 1995; Talleur et al., 2003; Taylor et al., 1997), and FAA regulations (FAA, 1997). This conclusion may help explain the small effects of using PCATDs for practice reported by Schneider et al. (2001). They chose to investigate the training of advanced VFR tasks. Such tasks are quite different from those of radio navigation, both in terms of the sensorimotor cues required and ability of a desktop simulator to support them.

Third, this device was seen as being of very little use for tasks from Primary Flight Training that required hovering as a part of the flight maneuver. Experts and students alike rated hovering tasks as unsupported or slightly supported. Hovering is an integral part of Primary Flight Training in IERW. Many tasks require achieving a stable hover, taking off from a hover, or landing to a hover. No evaluator was able to achieve a stable hover with the micro-simulator. The frustration was most noticeable with the high-time aviators who often refused to believe that they could “not make this thing hover.” The device simply did not allow pilots, all of whom had proven repeatedly that they could hover the aircraft, to meet performance standards. The stated reasons for this were legion: lack of peripheral visual cues, lack of visual cues to depth, lack of correct flight control response, lack of a helicopter flight model, even lack of platform motion cues. Unfortunately, the verdict on hovering the device gets worse upon closer examination. Those expert aviators who were best able to control heading, height above ground, pitch, and drift while attempting to hover stated that what they were doing with the flight controls was in no way similar to how they would operate the flight controls in a helicopter. In other words, the likelihood of positive transfer from hovering the device to hovering the aircraft is doubtful. This result is unique, to date, because no other published research has been found that combined PCATDs with helicopter flight-training tasks.

Fourth, the results from the answers to the general questions were consistent with earlier research. Nine of the 10 student evaluators (90%) had access to a PC at their place of residence. The result from an earlier assessment of Army IERW students was that 88% had access to a PC at their place of residence (S. Crouch, personal communication, February 8, 2001). The comparable data from Navy beginning flight students was 71% (Dunlap & Tarr, 1999). An Army-wide sample (ARI, 2001) found that 93% of all officers and 79% of all enlisted personnel had access to a PC at their place of residence. Four students (40%) had run a desktop flight simulator or aviation-related game prior to the evaluation. The result of an earlier Army IERW assessment was 34% (S. Crouch, personal communication, February 8, 2001). The figure for the Navy's beginning flight students was 47% (Dunlap & Tarr, 1999). At the very least, these figures show that beginning flight students are computer literate and that a substantial minority are using PCATDs.

Conclusions

Both highly experienced and student pilots evaluated a PCATD running Microsoft Flight Simulator 2000™ software. All conclusions are based on the specific hardware and software tested.

The high-time aviators and the student aviators agreed very closely in their evaluations of the utility of the micro-simulator for IERW training. The micro-simulator being evaluated, was rated as not able to support hovering tasks, and slightly able to support Primary Flight Training, especially upper air tasks. However, evaluators rated it as capable of supporting Instrument Flight Training (moderately to well), especially radio navigation tasks. Evaluators reported, in answer to a specific question, as well as in their spontaneous comments, that the simulator had promise as a procedures trainer for the instruments portion of IERW. Some also reported that it had value as a dynamic classroom aid for ground school instruction and as a practice device.

References

ARI (2001). *Internet access and PC capabilities.* (*Survey Rep.* No. 2001-08). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bell, H. H., & Waag, W. L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. *International Journal of Aviation Psychology*, *8*(3), 223-242.

Brannick, M. T., Prince, C., & Salas, E. (2005). Can PC-based systems enhance teamwork in the cockpit? *International Journal of Aviation Psychology*, *15*, 173-187.

Dennis, K. A., & Harris, D. (1998). Computer-based simulation as an adjunct to ab initio flight training. *International Journal of Aviation Psychology*, *8*(3), 261-276.

Dunlap, S., & Tarr, R. (1999). Micro-simulator systems for immersive learning environments. *Proceedings of the 21st Interservice/Industry Training Simulation and Education Conference* (pp. 506-511). Orlando, FL: I/ITSEC.

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

Jentsch, F., & Bowers, C. A. (1998). Evidence for the validity of PC-based simulators in studying aircrew coordination. *International Journal of Aviation Psychology*, *8*(3), 243-260.

Koonce, J. M., & Bramble, W. J. (1998). Personal computer-based flight training devices. *International Journal of Aviation Psychology*, *8*(3), 277-292.

Looking Glass Technologies (1995). Flight unlimited (computer software manual). Cambridge, MA: Author

Nielsen, J. (1993). *Usability engineering*. New York: Academic Press.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Norman, D. A. (1998). *The invisible computer*. Cambridge: The MIT Press.

Ortiz, G. A., Kopp, D., & Willenbucher, T. (1995). Instrument training using a computer-based simulation device at Lufthansa Airlines. In R.S. Jensen & L. A. Rakovan (Eds.). *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 1047-1050). Columbus, OH: The Ohio State University.

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.

Proctor, M. D., Panko, M., & Donovan, S. J. (2004). Considerations for training team situation awareness and task performance through PC-gamer simulated multiship helicopter operations. *International Journal of Aviation Psychology*, *14*(2), 191-205.

Roessingh, J. J. M. (2005). Transfer of manual flying skills from PC-based simulation to actual flight-comparison of in-flight measured data and instructor ratings. *International Journal of Aviation Psychology*, *15,* 67-90.

Schneider, D., Greene, R., Levi, K., & Jeffery, D. S. (2001). Microflight simulator. *Proceedings of the 23rd Interservice/Industry Training Simulation and Education Conference* (pp. 1319-1325). Orlando, FL: I/ITSEC.

Talleur, D. A., Taylor, H. L., Emanuel, T. W., Rantanen, E. M., & Bradshaw, G. L. (2003). Personal computer aviation training devices: Their effectiveness for maintaining instrument currency. *International Journal of Aviation Psychology*, *13*(4), 387-399.

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D., Emanuel, T., & Phillips, S. (1997). *Transfer of training effectiveness of personal computer-based aviation training devices.* (DOT/FAA/AM-97/11). Washington, DC: Office of Aviation Medicine.

Taylor, H. L., Talleur, D. A., Emanuel, T. W., & Rantanen, E. M. (2004). *The effectiveness of personal computer aviation training devices (PCATDs), flight training devices (FTDs), and an airplane in conducting instrument proficiency checks*. Paper presented at the Annual Midyear Symposium of APA Division 19 and 21 and the Potomac Chapter of the HFES, Ft. Belvoir, VA.

Taylor, H. L., Talleur, D. A., Rantanen, E. M., & Emanuel, T. W. (2005). *The effectiveness of a personal computer aviation training device (PCATD), a flight training device (FTD), and an airplane in conducting instrument* *proficiency checks*. Paper presented at the 13th International Symposium on Aviation Psychology, Dayton, OH.

U. S. Army Aviation Center (November, 1999). *Initial entry rotary wing aviator course: Primary phase*. Fort Rucker, Alabama: Author.

U. S. Army Aviation Center (February, 2001*). Initial entry rotary wing aviator course: Instrument phase*. Fort Rucker, Alabama: Author.

International Journal of Applied Aviation Studies, 5, Number 2

Transfer of Learning Between a Small Technically Advanced Aircraft and a Commercial Jet Transport Simulator

Stephen M. Casner
National Aeronautics and Space Administration
Ames Research Center
Mail Stop 262-4
Moffett Field, CA 94035-1000

Abstract

This study examines the extent to which skills acquired using advanced avionics found in a small technically advanced aircraft (TAA) transfer to more sophisticated equipment found in a modern jet transport. Eight pilots were trained to proficiency on twenty-eight procedures using the GPS navigation computer, autopilot, and flight director system found in a small technically advanced aircraft. Eight other pilots did not receive the training. All sixteen pilots were then tested on their ability to perform the same procedures using a computer-based simulator of the flight management and guidance systems found in a popular jet transport. Pilots attempted the jet transport procedures with no prior exposure to the equipment, no training, and no reference materials. Pilots who received the TAA training successfully completed 83% of all procedures in the jet transport, while pilots in the control group achieved an average success rate of 54%. Further analysis of the data showed that much of the control group's success was attributable to superficial strategies guided by labels that appear on the knobs and buttons of the equipment, and that their scores averaged only 22% on procedures for which no label cues were available. The results cast a strong vote for transfer of learning between the two types of equipment, and for the use of small technically advanced aircraft to train pilots who will later transition to the commercial jet fleet.

Introduction

Among the challenges of transitioning from small piston training airplanes to the commercial jet fleet is the requirement of learning to use the advanced avionics systems found in the modern commercial jet cockpit. Commercial air carriers have long struggled with training pilots who transition from the general aviation environment, or from other non-glass cockpit equipped aircraft (Wiener, 1985; Sarter & Woods, 1995; Palmer, Hutchins, Ritter, & van Cleemput, 1993; FAA, 1996; Risukhin, 2001; Billings, 1997; ATA, 1997; ATA, 1998, ATA, 1999).

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

Many of the advanced avionics systems found in commercial jet aircraft are now becoming widely available in small general aviation aircraft. The current generation of small *technically advanced aircraft* (TAA) offers less-sophisticated versions of the same equipment found in the commercial jet fleet. Figures 1 and 2 offer a simple comparison between the capabilities of current-generation commercial jets and small technically advanced aircraft. Both airplanes offer the same basic configuration of navigation computer, multifunction displays, flight director, and autopilot systems.

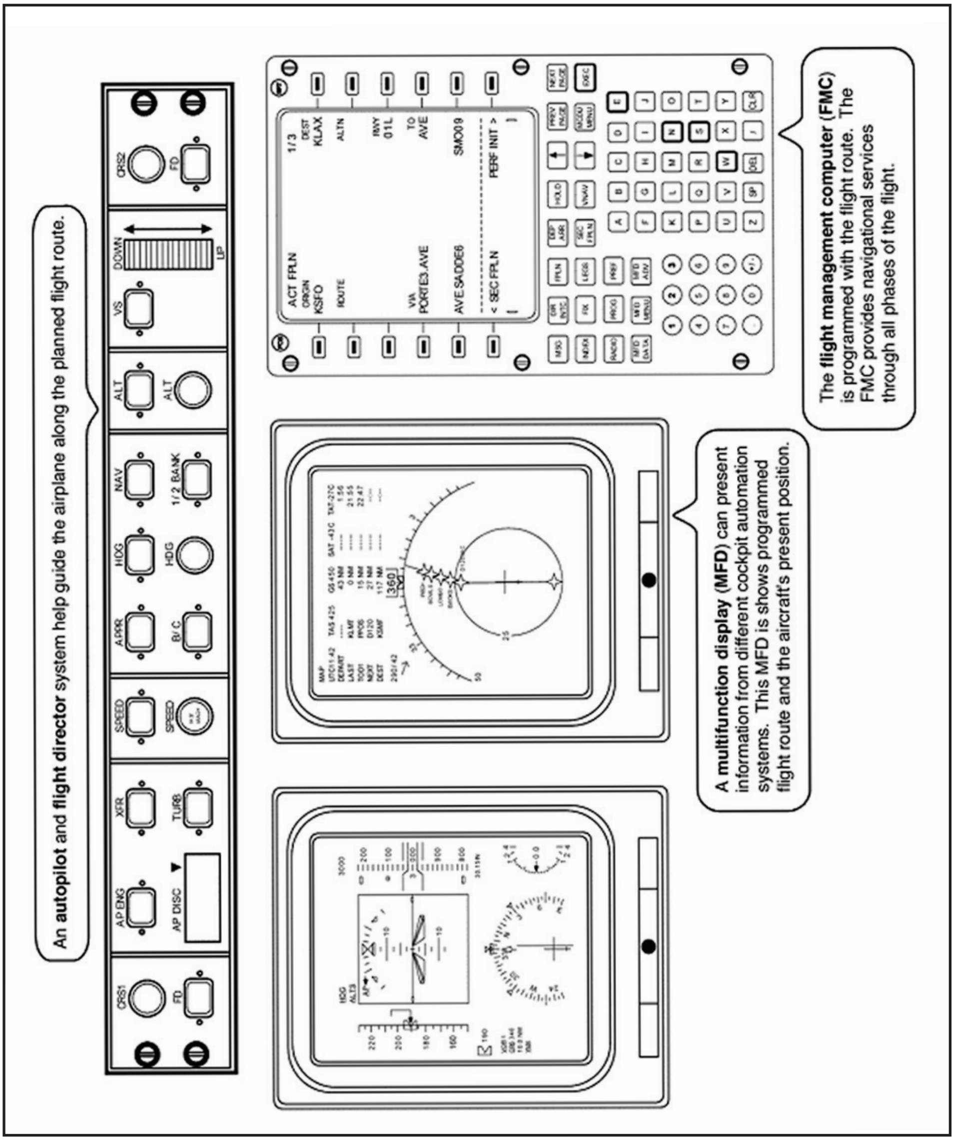


Figure 1.

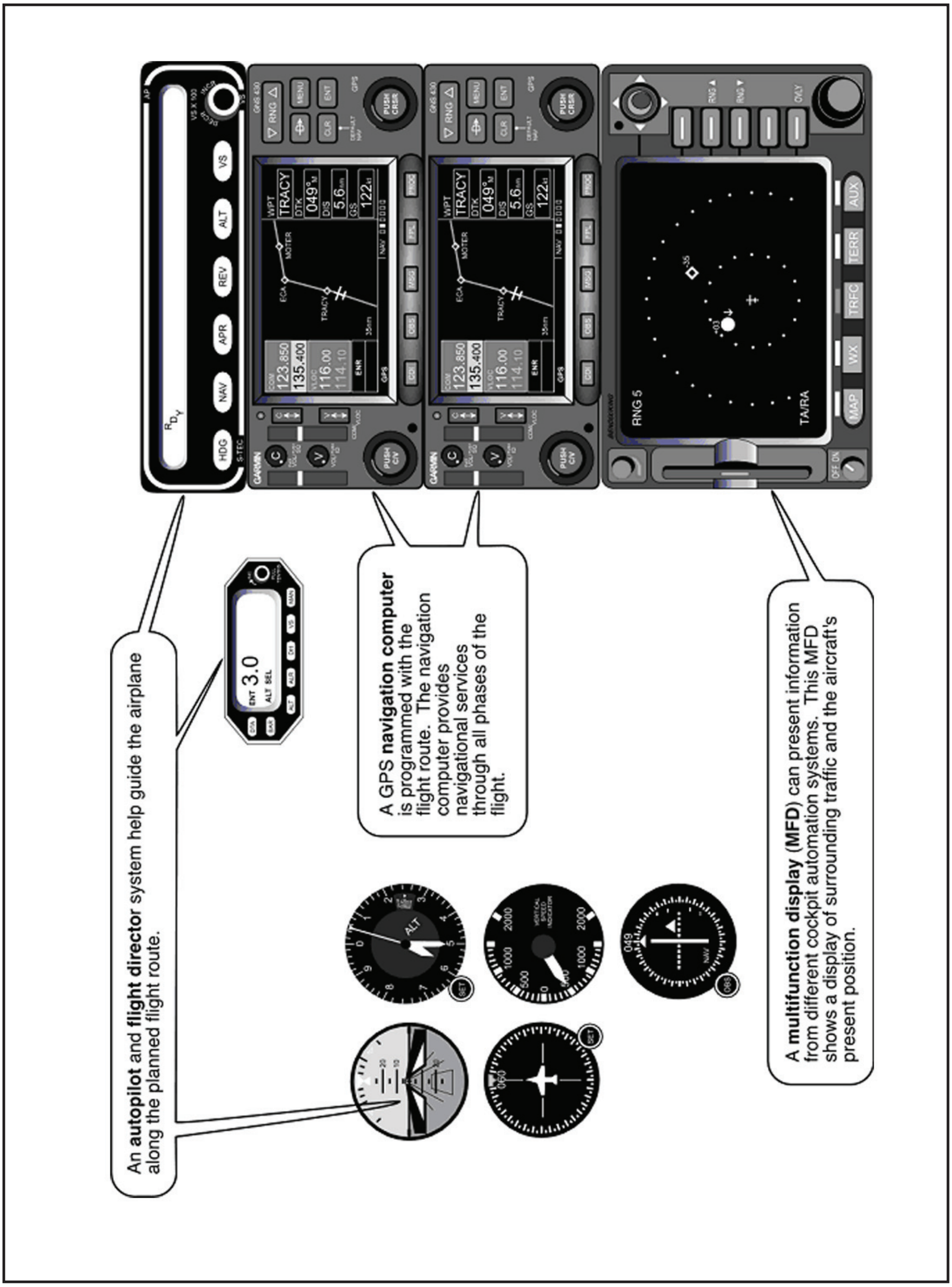

An autopilot and flight director system help guide the airplane along the planned flight route.
A GPS navigation computer is programmed with the flight route. The navigation computer provides navigational services through all phases of the flight.
A multifunction display (MFD) can present information from different cockpit automation systems. This MFD shows a display of surrounding traffic and the aircraft's present position.

Figure 2.

The similarity between the equipment found in both types of aircraft raises a practical question: To what extent do concepts and skills learned in a small technically advanced aircraft transfer to the more sophisticated equipment found in the modern jet transport?

This study aims to answer this question about knowledge and skill transfer in a practical way. Eight instrument-rated pilots with no previous experience with advanced avionics were trained to proficiency on twenty-eight procedures in a small technically advanced aircraft. These procedures were derived from the FAA Instrument Rating Practical Test Standards (FAA, 2004). Upon completion of this training, pilots were asked to perform the same procedures using a computer-based simulation of a popular jet transport airplane. A control group consisting of eight pilots who did not receive the TAA training was also asked to perform the procedures in the commercial jet simulator.

Table 1 shows the procedures that pilots were asked to master in the technically advanced aircraft, and later asked to perform in the commercial jet simulator.

Table 1
Procedures learned in the TAA and then tested using the jet simulator.

PROCEDURES	SESSIONS
Navigation Computer	
Access page	1, 2, 3, 4, 5
Find information on page	1, 2, 3, 4, 5
Simple data entry	1, 2, 3, 4, 5
Access extended page	1, 2, 3, 4, 5
Check navigation database	1, 2, 3, 4, 5
Database lookup	1, 2, 3, 4, 5
Enter origin and destination	1, 2, 3, 4, 5
Menu select procedures	1, 2, 3, 4, 5
Review route	1, 2, 3, 4, 5
Set active waypoint	2, 3, 4, 5
Set inbound course	2, 3, 4, 5
En Route	
Announce active waypoint	1, 2, 3, 4, 5
Find time and distance to active waypoint	1, 2, 3, 4, 5
En Route Modifications	
Direct to	2, 3, 4, 5
Add waypoint	1
Delete waypoint	1

Multiple holds	3, 4, 5
Program hold	3, 4, 5
Exit hold	3, 4, 5
Program crossing restriction	2, 4, 5
Autopilot	
Altitude pre-select	4, 5
Engage Vertical Speed function	4, 5
Determine necessary vertical speed	4, 5
Fly heading using Heading Select	4, 5
Verify engaged mode	4, 5
Arm Nav function	4, 5
Verify Nav armed	4, 5
Constant speed descent	4, 5

Method

Participants

Sixteen commercial instrument-rated pilots were recruited from several local flight schools. Pilots ranged from 300 to 1,600 hours of flight experience with a mean of 1,106 hours. Pilots were told they would not be paid for their participation but would receive instrument flight experience or simulated flight experience using advanced avionics.

Procedure

The sixteen pilots were divided randomly into two groups. Pilots in the TAA group were trained to perform the twenty-eight procedures in small technically advanced aircraft, and were then asked to perform the same procedures using the computer-based simulator of the jet transport.

Pilots in the control group were asked to perform all of the procedures using the computer-based simulator, without receiving the technically advanced aircraft training.

One purpose of the control group was to factor out any successes that might be enjoyed due to what Irving, Polson, and Irving (1994) referred to as *label following*. Label following occurs when a computer system provides simple cues about how it might be operated, typically in the form of labels that suggest the purpose or operation of knobs, buttons, and dials on the equipment. When using label following, operators can often succeed in completing a task without any knowledge or skill related to that task. For example, consider the task of calling up the Index page on the control display unit (CDU) shown in Figure 3. A person with little or no knowledge about this system might notice the button labeled INDEX, shown on CDU, and correctly hypothesize that pushing this button will accomplish the task.

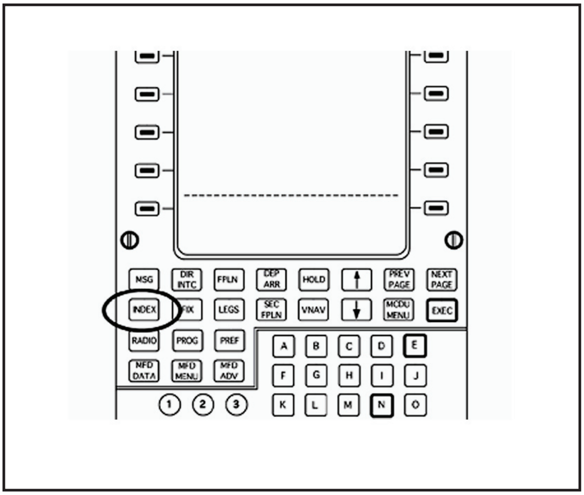


Figure 3.

Although label following cues are legitimate components of expert performance, we would like to distinguish between success attributable to true understanding of the system and tasks, and success due to label following.

Technically advanced aircraft training. For the eight pilots participating in the TAA training group, the technically advanced aircraft training occurred in five scheduled sessions that covered the 28 procedures listed in Table 1.

Prior to each session, each pilot was told to read a technical publication (Casner, 2002) that described the concepts and skills required to perform any new procedures that would be covered during that session. Pilots were told to master the material as best as they could, and that during the upcoming session, they would have the opportunity to demonstrate and practice their newly learned procedures in flight. Pilots were told that they would also be given an opportunity to practice all of the procedures that they had learned during the previous sessions. The second column in Table 1 shows the sessions in which each procedure was practiced. Note that some procedures were practiced only during one or two sessions during the course of the training. Note also that no new procedures were introduced during the fifth session as this session was intended as a “check flight” to ensure that all pilots understood and were able to perform all procedures.

During each session, the experimenter briefly reviewed the procedures that would be practiced during the flight, provided the pilot with charts covering the routes and approaches to be flown, and answered any questions the pilot had about the reading.

During each dual-instruction flight, the experimenter rode in the right seat but did not operate the controls. A script for each flight was prepared in advance and used by the experimenter to ensure that each pilot was presented with the same procedures in the same order.

The jet transport simulator evaluation. Following the conclusion of the TAA training sessions, all sixteen pilots participated in a session in which they were asked to perform the same 28 procedures shown in Table 1 using a computer-based simulation of the cockpit of a popular jet transport airplane. Again, eight of the pilots had received the technically advanced aircraft training and eight had not. It was explained to all sixteen pilots that no training on the jet transport systems would be provided, nor would pilots have the opportunity to access any reference materials for the systems. The aim of the study was to determine to what extent pilots' existing knowledge could help guide them through the procedures. The TAA training group had their instrument flying skills together with their technically advanced aircraft training. The control group had their instrument flying skills to guide them, together with any label following cues present on the jet transport equipment.

During the jet transport simulator session, pilots were presented with procedures and asked to do their best to perform them without asking for intervention from the experimenter. If the pilot was able to successfully complete a procedure, a score of 1 was recorded for that procedure. If an error was made, a score of 0 was recorded for the procedure. If an impasse was encountered, pilots could ask for intervention, these interventions were recorded, and a score of 0 was recorded for the procedure. If a pilot was unsuccessful on a particular procedure, the experimenter demonstrated the procedure before moving on to the next. Since the jet transport travels as much as five times faster than the piston airplane, the simulation was frozen while the experimenter took the time to provide the needed interventions. A paper scorecard was used to record scores for each procedure.

Results and Discussion

Overall Performance

A first question posed by the experiment is the extent to which performance for procedures performed in the jet simulator was leveraged by the technically advanced aircraft training. Figure 4 shows a graph of the proportion of procedures successfully completed by each pilot using the jet simulator. The data in Figure 4b represent individual scores (proportions for all 28 procedures combined) for the sixteen pilots. The pilots who received the technically advanced aircraft training performed significantly better than the control group ($df = 14$, $t = 7.72$, $p < .0001$).

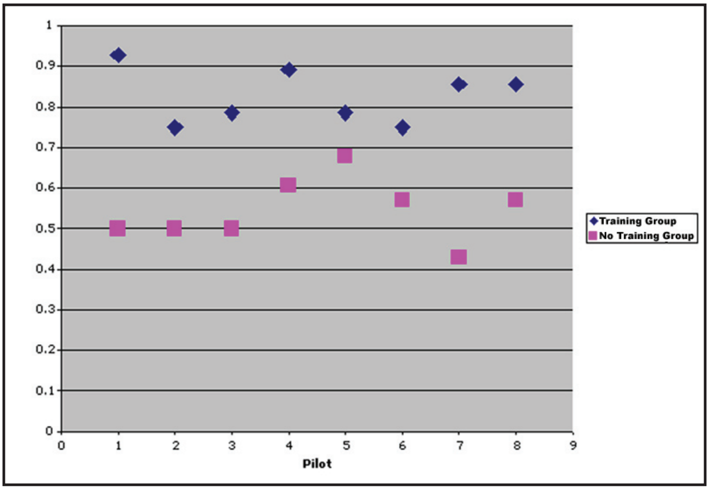


Figure 4. Individual scores for both groups for the jet simulator procedures.

The overall performance of the TAA training group casts a vote for the transfer of knowledge and skill from the small technically advanced aircraft to the jet. These pilots were able to successfully perform 83% of all procedures on the jet on the first try.

Success Due To Label Following

The mean success rate of 54% for the control group prompts the question of to what extent their success was attributable to label following. To answer this question, procedures were segregated into two categories, those for which label cues appeared on the equipment, and those for which no cues appeared. The graphs in Figures 5a and 5b show the results for the TAA training and control groups on label-cued and non-label-cued procedures.

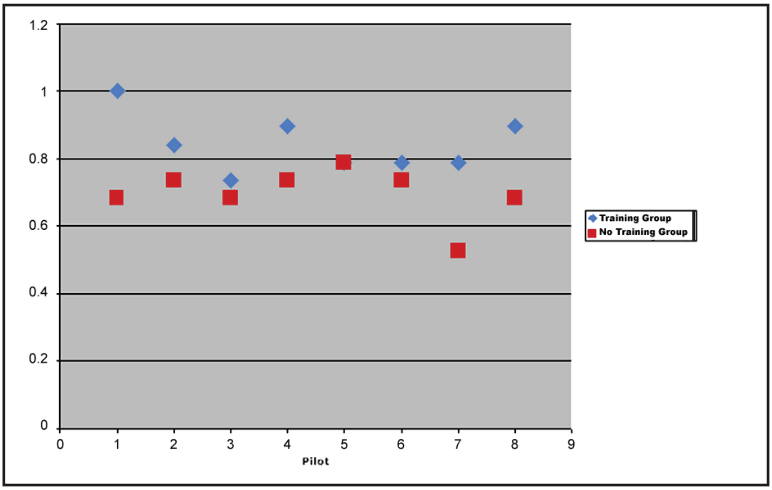


Figure 5a. Individual scores for procedures with label cues.

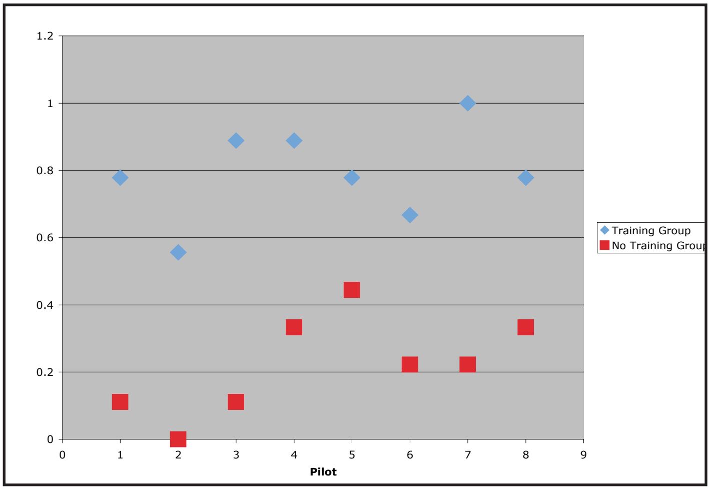


Figure 5b. Individual scores for procedures without label cues.

A 2-way analysis of variance (ANOVA) reiterated the main effect of the advantage due to receiving the TAA training ($F=76.3$, $p < .0001$), a main effect due to the presence of label cues ($F=41.2$, $p < .0001$), and a significant interaction between the two factors ($F=27.0$, $p < .0001$).

For the pilots who received the TAA training, the same ANOVA revealed no significant effect due to presence or absence of label cues, suggesting that the cues provided by pilots' own knowledge were as strong as the cues provided by the labels. The pilots who did not receive the TAA training performed well when label cues were present but poorly in the absence of label cues. This suggested that their successes occurred in the absence of understanding of how to operate the systems. Lastly, the pilots who received the TAA training performed significantly better than their control group counterparts on tasks for which label following was possible. This suggested that the TAA training group imparted knowledge on tasks even when label cues were present, and this knowledge led to significantly better performance.

Breakdown By Procedure

Table 2 shows the results (across all pilots) for the twenty-eight individual procedures on which all pilots were tested.

Table 2

Average scores for the two groups on each of the 28 procedures tested in the jet simulator.

TASKS PERFORMED USING THE JET SIMULATOR	Labels	Control Group	Trained Group	t-test
Navigation Computer				
Access page	L	1	1	
Find information on page	**L**	1	1	

Table is continued on following page

Simple data entry	**L**	0.75	1	
Access extended page		0.375	0.75	
Check navigation database	**L**	0.25	0.875	p < .01
Database lookup	**L**	0.625	0.625	
Enter origin and destination	**L**	0.75	0.875	
Menu select procedures	**L**	0.875	1	
Review route	**L**	0.875	1	
Set active waypoint		0.25	0.875	p < .01
Set inbound course		0.125	0.5	
En Route				
Announce active waypoint	**L**	**1**	**1**	
Find time and distance to active waypoint	**L**	0.875	1	
En Route Modifications				
Direct to	**L**	0.75	0.875	
Add waypoint		0.125	0.75	p < .01
Delete waypoint		0.25	0.875	p < .01
Multiple holds	**L**	0.375	1	p < .01
Program hold	**L**	0.25	0.5	
Exit hold	**L**	1	0.875	
Program crossing restriction	**L**	0.5	0.5	
Autopilot				
Altitude pre-select	**L**	0.125	0.25	
Engage Vertical Speed function		0.125	0.875	p < .01
Determine necessary vertical speed		0.375	0.875	p < .05
Fly heading using Heading Select	**L**	0.75	0.75	
Verify engaged mode	**L**	0.5	0.875	
Arm Nav function		0.375	0.875	p < .05
Verify Nav armed		0	0.75	p < .01
Constant speed descent	**L**	1	1	

Our hypothesis was that pilots who learned the 28 procedures during the TAA training would perform well on the same procedures on the jet transport simulator. The data in Table 2 show that pilots in the TAA training group enjoyed at least a 50% success rate on 27 of the 28 procedures, and at least a 75% success rate on 23 of the 28 procedures. Overall, the results cast another vote of confidence for the hypothesis that the technically advanced aircraft training positively transferred to the jet simulator.

Pilots who received technically advanced aircraft training performed significantly better than the control group on nine of the 28 procedures. Among these nine procedures, seven were procedures for which no label cues were available. This result strongly suggests that pilots' success for these procedures was derived from concepts and skills they had learned during the technically advanced aircraft training.

It is also interesting to note that the TAA training failed to leverage pilots' performance on the Altitude preselect task, even though both the concepts and button-pushing steps for this procedure are quite similar in the two aircraft.

The following six procedures are required to perform the "course intercept" maneuver required by the instrument rating practical test standards: Fly heading using Heading Select, Verify engaged mode, Set active waypoint, Set inbound course, Arm Nav function and Verify Nav armed. Previous studies with experienced airline pilots have demonstrated the difficulty that pilots encounter when learning to perform the course intercept maneuver using advanced avionics (Irving, Polson, & Irving, 1994). The course intercept procedure combines several advanced concepts such as the notions of departing and rejoining the planned route, and armed vs. engaged autopilot modes. Slightly fewer than 70% of Irving et al's airline pilots who had just completed an airline initial training course on a Boeing 737 were able to successfully complete this procedure following explicit training using the same equipment used for the test. Pilots who completed the technically advanced aircraft training completed this collection of procedures successfully 77% of the time, using equipment they had never seen before. The control group was successful 33% of the time.

It was noted earlier that some procedures were only practiced once or twice during the course of the TAA training. The aggregate success rate for these procedures was 75%, while the mean for the remaining procedures, which were practiced either four or time times during the course of the training, was 83%. There was no significant difference between these means.

Correlating Total Flight Time and Performance with Cockpit Automation

A last interesting analysis is to look at the relationship between the total flight experience of each pilot and their scores for the jet transport procedures. For the group that received the TAA training, the correlation was –0.43. For the group that did not receive the TAA training, the correlation was 0.028. These results agreed with the findings of Casner (2004) and suggested that total flight experience alone provides little or no intuitive guidance for the use of advanced avionics, nor can it serve as a substitute for learning and experience with advanced avionics. Advanced avionics proficiency appears to be a unique set of skills that must be learned in addition to basic airmanship.

Conclusion

The results suggested that time invested in mastering skills in a small technically advanced aircraft can have a significant impact on the subsequent learning of more sophisticated equipment. The systems now found in small technically advanced training airplanes appear to provide a simple, cost-effective way of introducing advanced avionics to pilots who are still in the formative phases of their professional aviation careers. This should alleviate the problem of new-hire pilots arriving to airline initial training programs with little or no experience in technically advanced cockpits.

We should be careful not to interpret pilots' performance with the jet simulator too literally. Pilots' performance using the jet simulator was scored on a task-by-task basis. Correctly performing a large proportion of tasks using a simulator does not necessarily add up to real-time successful operation of a jet aircraft in a full-mission flight environment. It is well known that the concurrent performance of the many individual tasks that make up the job of piloting an aircraft is a skill of its own: one that requires additional practice once the individual skills have been mastered. The most reasonable interpretation of the results presented here is that the small airplane training and experience places the pilot-in-training farther along in the learning process: shortening the time and effort required to train the future jet transport pilot.

A second issue, not directly addressed by this study, is the importance of learning underlying principles about the systems found in technically advanced aircraft, in addition to learning button-pushing procedures. Previous studies have demonstrated that learning focused on knobs, dials, and procedures results in fast training times, but also tends to result in brittle skills. These brittle skills are typically not transferable to other equipment, or problems and situations that are different from those initially learned. Kieras and Bovair (1984) demonstrated how students who received "how it works" explanations for a set of procedures they had learned were significantly more successful when presented with related problems that challenged them in different ways. Pennington, Nicolich, and Rahm (1995) conducted a similar study. Recognizing that conceptual learning can happen even when explicit conceptual instruction is not given, Chi, Bassok, Lewis, Reimann, and Glaser (1989) demonstrated how students generate and successfully apply their own "self-explanations" while solving problems.

Looking at Figures 1 and 2, we can see that the knobs, dials, and procedures used to operate the systems found in each airplane are different. The success of the group who received the TAA training can only be attributed to pilots' acquisition and application of generalized concepts and principles.

A future study could directly address the question of learning *generalizable* concepts and skills in three ways: (1) by measuring the effects of different amounts of conceptual explanations provided to pilots; (2) by examining the self-explanations generated by pilots in the absence of explicit conceptual instruction; and (3) by measuring pilots' ability to perform procedures that were not covered during training.

References

Air Transport Association (1997). *Towards an operational philosophy and model training program for FMS-generation aircraft.* First report of the ATA Human Factors Committee, Automation Subcommittee.

Air Transport Association (1998). *Potential knowledge, policy, or training gaps regarding operation of FMS-generation aircraft.* Second report of the ATA Human Factors Committee, Automation Subcommittee

Air Transport Association (1999). *Performance of standard navigation tasks by FMS-generation aircraft.* Third report of the ATA Human Factors Committee, Automation Subcommittee.

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

Casner, S. M. (2002). *Cockpit automation for general aviators and future airline pilots.* Ames, Iowa: Blackwell Publishing.

Casner, S. M. (2004). Flying IFR with GPS: How much practice is needed? *International Journal of Applied Aviation Studies*. *4*(2), 81-97.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science 13*, 145-182.

Federal Aviation Administration (2004). Instrument rating practical test standards. FAA-S-8081-4D.

Federal Aviation Administration Human Factors Team (1996). *Report on the interfaces between flightcrews and modern flight deck systems.* Washington, D. C.: U.S. Department of Transportation, Federal Aviation Administration, June 18, 1996.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science 8*, 255-273.

Irving, S., Polson, P. G., & Irving, J. E. (1994). A GOMS analysis of the advanced automated cockpit. *Proceedings of CHI '94: Human factors in computer systems*, New York: Association for Computing Machinery, 344-350.

Palmer, E. A., Hutchins, E. L., Ritter, R. D., & van Cleemput, I. M. (1993). *Altitude deviations: Breakdowns of an error-tolerant system,* NASA Technical Memorandum.

Pennington, N., Nicolich, R., & Rahm, J. (1995). Transfer of training between cognitive subskills: Is knowledge use specific? *Cognitive Psychology 28*, 175-224.

Risukhin, V. (2001). *Controlling pilot error: Automation*. New York: McGraw-Hill.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors 37*(1), 5-19.

Wiener, E. L. (1985). *Human factors of cockpit automation: A field study of flight crew transition* (Technical Report: 118). National Aeronautics and Space Administration.

International Journal of Applied Aviation Studies, 5, Number 2

Pilots' Perspectives on the Pilot-monitored Approach: Findings from a Web Survey

Kathryn Wochinger and Deborah A. Boehm-Davis
George Mason University
Fairfax, VA

Correspondence Information:

Deborah A. Boehm-Davis
George Mason University
MSN 3F5
Fairfax, VA 22030-4444
dbdavis@gmu.edu
703-993-8735

Kathryn Wochinger
Mitretek Systems
955 L'Enfant Plaza North, SW
Suite 1000
Washington, D.C. 20024
kathryn.wochinger@mitretek.org
202-554-8107

Abstract

A type of instrumented approach referred to as the pilot-monitored approach may have safety advantages over standard instrumented approaches, but there are insufficient data to determine the extent of its value. Likewise, there are no federal standards or guidelines in regards to its training or procedures, nor is information readily available on its practice. However, the catastrophic crash of Korean Air Flight 801 highlighted a need for research on the pilot-monitored approach. This paper presents the results of an initial research effort designed to obtain basic information on issues such as the frequency of its use and its perceived value. The research used a questionnaire posted on the World Wide Web to survey professional pilots about their experiences with and knowledge of the pilot-monitored approach, as well as their assessments of its safety and utility. The analysis focused on responses from 205 pilots who held either a commercial or an airline transport certificate. Results suggested that many pilots view the pilot-monitored approach as a safe and

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

valuable means of executing instrumented approaches. However, the findings also showed that there is much uncertainty about very basic aspects of the pilot-monitored approach, including its definition, role in aircraft with auto-land capability, and the existence of federal policy toward it. The results suggested that further investigation is warranted to determine whether there is a need to standardize training and/or procedures for the pilot-monitored approach.

Pilots' Perspectives on the Pilot-monitored Approach: Findings from a Web Survey

The pilot-monitored approach is a means of executing an instrumented approach in which one pilot focuses on the cockpit instruments that display the aircraft's deviation from its intended path, while the other pilot scans the external environment for signs of the runway. The approach is designed to be executed when the runway environment is obscured by poor visibility conditions and it is desirable or necessary to provide resources for the crew to acquire visual reference to the runway. Currently, there is relatively little information available on the pilot-monitored approach, including basic issues such as the frequency of its use, criteria for training, and pilot opinion of its safety and usefulness. However, the catastrophic crash of Korean Air (KAL) Flight 801 in 1997 highlighted a need to improve understanding of the pilot-monitored approach.

KAL Flight 801, operating in United States airspace as a regularly scheduled international passenger service flight, was attempting to land at Guam International Airport when it crashed into high terrain three miles short of the runway. The approach was executed under difficult circumstances, at night with rainy weather and a fatigued captain. The crew was executing the Instrument Landing System (ILS) approach to runway 6 Left. The ILS is a precision approach and landing aid that consists of a localizer, a glide slope, marker beacons and an approach light system (Nolan, 2004). The glide slope provides vertical guidance and the localizer provides lateral guidance to the pilot in lining up the plane with the runway. In this case, however, the glide slope portion of the ILS on runway 6 Left was out of service for reconstruction (National Transportation Safety Board, 2000). Shortly after being cleared to land, the airplane crashed into hilly terrain. The crash and a post crash fire destroyed the aircraft and killed 228 out of the 254 people on board.

The NTSB attributed the probable cause of the accident to the captain's failure to adequately brief and execute the nonprecision approach, as well as the first officer's failure to monitor and cross-check the captain's execution of the approach (NTSB, 2000). The NTSB noted that the flight crew did not use the pilot-monitored approach. Rather, the Captain handled the flight controls and assumed responsibility for scanning the external view during the approach and final descent. Among the dozens of recommendations issued by the NTSB as a result of its investigations was one to, "Conduct or sponsor research to determine the most effective use of the monitored approach and the maximum degree to which it can be safely used and then require air carriers to modify their procedures accordingly" (p. 148).

Description

The pilot-monitored approach is designed to be used in poor visibility conditions, which intensify the difficulty in switching visual attention between the instrument landing display and the runway environment. Further, weather conditions that frequently accompany poor visibility, such as surface winds, rain, and/or snow, are likely to increase pilot workload during the approach and landing. The monitored approach allows the landing pilot to focus on the external runway environment by freeing him or her from the task of following the instruments. This objective is met by tasking one pilot (the Non-Landing Pilot) with following instruments until decision height (the flight altitude at which a landing decision needs to be made), and tasking the other pilot (the Landing Pilot) with acquiring visual signs of the runway. The Landing Pilot, who has maintained visual reference of the runway, makes the decision whether to land or execute a go-around procedure. In the event of a landing, the Landing Pilot assumes control of the airplane and guides the aircraft to the runway. In the event of a go-around procedure, the Non-Landing Pilot maintains control and executes a go-around procedure. Frequently, the captain is the Landing Pilot and the Non-Landing Pilot is the first officer. The pilot-monitored approach contrasts with standard means of executing approaches in which the Landing Pilot scans both the flight controls and the external runway environment, and maintains control of the aircraft during both the approach and landing.

Scarce Information Available

A literature review revealed little data on the pilot-monitored approach, although there is some information available in Federal Aviation Administration (FAA) Advisory Circulars, Aviation Regulations, and the Aeronautical Information Manual (AIM). However, these documents do not directly address the pilot-monitored approach nor define official FAA policy toward it. Potential sources of information are the training manuals and standard operating procedures of air carriers. However, these sources are proprietary and their acquisition is difficult. Because the procedure has inherent room for variations, such as in the number of altitude call-outs or communication between the pilots, it is possible that carriers vary in their procedures or policies in regards to the pilot-monitored approach. Unfortunately, this information is not readily available.

Post-crash reports may be the best available source of information. For example, the report of the NTSB investigation into the crash of American Airlines Flight 1340 in Chicago in 1998 describes the policy of American Airlines for Category (CAT) II approaches (NTSB, 2001). According to the report, the American Airlines 727 Operating Manual states that CAT II approaches are to be flown with the autopilot coupled to the ILS until decision height. The CAT II procedure requires the First Officer to take the role of the Flying (Non-Landing) Pilot who "remains on instruments throughout the approach and landing, and makes all normal callouts below 1,000 feet" (p. 22). When the Captain visually acquires the runway and is ready to take control of the plane and complete the approach and landing visually, he or she will "push the first officer's hand from the throttles and call out 'I've got it', indicating intention to land" (p. 22).

Research Questions

Due to the scarcity of information available on the pilot-monitored approach, the survey research reported below was an initial effort to obtain information from pilots about their experiences and opinions on the monitored approach. The survey was designed to address basic but key issues, including:

- Pilot opinion of the safety and utility of the pilot-monitored approach,
- Pilot experiences using the pilot-monitored approach,
- Pilot recommendations about its improvement.

Method

The survey included closed and open-ended questions in regards to the pilot's familiarity with the pilot-monitored approach, knowledge of its procedures, and awareness of FAA policy towards it, as well as its perceived safety, difficulty, and training requirements. The survey included open-ended questions about its strengths and weaknesses, and areas for improvement. In this way, the survey served as a means for pilots to voice their opinions and describe their personal experiences in regards to the monitored approach. The research team developed the survey with the help of several professional pilots, both active and retired, and piloted the survey with several professional pilots.

Recruiting participants for the survey consisted of two parts. First, recruitment letters were mailed to twelve pilot unions, predominantly of the major carriers, and about sixty non-union pilot organizations, flight schools and university aviation departments informing them of the survey's existence and encouraging their pilot members to complete it. Second, notices were posted on internet pilot forums about the survey. In each of these cases, a code was provided that would allow pilots to enter the survey web site and submit their responses.

The survey was posted on the World Wide Web from April 2004 through September 2004. There were two rounds of data collection, the first from April to June, and the second from July to September. The second round had fewer survey items than the first because feedback suggested that the survey was somewhat too long.

Results

Respondents

The design of the web site required visitors to enter a code to navigate to the survey. Visitors who entered a code may have opted not to take the survey, or may have started to take the survey but not completed it. The analysis only used data from individuals who completed all or nearly all of the survey and whose comments revealed that they understood the definition of a pilot-monitored approach. Data from individuals who completed a page or two were not included in the analysis. In addition, surveys from individuals who revealed through their comments that they were referring to a type of monitored approach performed by Air Traffic Controllers were not included in the analysis. The final data set consists of 205 records.

Table 1 summarizes the pilots' background in terms of their airline certificates held and their years of experience. The majority of pilots (n=133) exclusively held airline transport airman certificates.

Table 1
Pilot Background in Terms of Airman Certificate and Years of Experience

Years Experience	Commercial	Airline Transport	Holds Both Certificates	Total
1-5 years	17	7	4	28
6-10 years	6	17	5	28
11-15 years	1	10	5	16
16-20 years	-	15	5	20
more than 20	4	83	8	96
Total	28	133	27	191*

*Fourteen pilots did not provide information about their certificates or years of experience.

Familiarity with the Term and Procedures

The first page of the survey focused on familiarity with the pilot-monitored approach in terms of its meaning, procedures, and FAA policy. Most of the respondents indicated that they were familiar with the term and its procedures. For example, in response to the question, "How would you rate your familiarity with the term monitored approach," 79% of respondents (n=162) said they are "pretty familiar" or "very familiar" with the term; 7.3% of respondents (n=15) said that they are "minimally" or "not at all familiar" and 13.7% (n=28) said they are "somewhat familiar." Similarly, more respondents agreed (77.6%, n=159) than disagreed (13.7%, n=28) with the statement, "In the aircraft that I have flown as a professional pilot, I had a good understanding of the procedures required to execute a monitored approach," and 8.8% (n=18) were not sure. Thus, the majority of pilots who completed the survey indicated that they were familiar with the term and the procedures of the pilot-monitored approach. This result is not very surprising considering that individuals not familiar with a topic may be unlikely to pursue a survey about it, and the findings should be interpreted in the context of this particular sample of pilots.

Confusion about FAA Policy

The results showed that there is uncertainty among the surveyed pilots about the existence of an official policy on the pilot-monitored approach (Note: there is no official FAA policy). Pilots were asked to agree or disagree with the statement, "The FAA requires IFR [instrument flight rules] pilots to follow the monitored approach method under some circumstances." The responses varied: 36.6% (n=75) said they were not sure, 37.5% (n=77) agreed with the statement, and 25.9% (n=53) disagreed with the statement. It is noteworthy that the sample of pilots, who for the most part described themselves as being familiar with the pilot-monitored approach, was uncertain about FAA policy. This finding suggested that there might be uncertainty about FAA policy in the wider aviation community.

Echoing this state of uncertainty is the fact that 49% (n=101) of the pilots agreed with the statement that, "The term monitored approach means different things to different pilots;" 25.9% (n=53) were not sure; and 23.4% (n=48) disagreed. Three respondents did not answer the question. One source of confusion lies in the fact that Air Traffic Control (ATC) can monitor an approach and inform the flight crew when the aircraft deviates from the approach track.

Several pilots commented on the need for a standardized procedure and a federal policy. For example, one pilot wrote, "I believe there should be a recommended standard monitored approach published as an Advisory Circular. This would encourage corporate flight departments to utilize the procedure."

Ambiguity of the Term

The ambiguity of the term is illustrated by the fact that 17 individuals who participated in the survey believed that the monitored approach referred to ATC monitoring aircraft. This fact became apparent from the comments that they submitted at the end of the survey, in which they discussed ATC procedures. In addition, pilots who were familiar with the pilot-monitored approach were not necessarily familiar with the term. For example, one pilot noted that he or she "wasn't aware that it was labeled as such" until taking the survey, yet this pilot's responses indicated that he or she had trained and executed the pilot-monitored approach. One pilot suggested that the term "Pilot-Monitored Approach" is "a better way" to describe the method than simply "monitored approach." It is possible that the term "pilot-monitored approach" would clarify its meaning by distinguishing it from ATC-monitored approach.

Current Use

An important but unanswered question concerns frequency of use. To help answer this question, the survey asked pilots to agree or disagree with the statement that, "Pilots rarely use the monitored approach method in commercial or business aviation;" 37.5% (n=77) agreed; 36.6% (n=75) were not sure; and 25.9% (n=53) disagreed. The range of responses mirror the ambiguity and uncertainty reflected in the distribution of responses to the question about FAA policy, suggesting that there is incomplete information about this method of precision approaches.

The survey asked pilots to indicate "the number of times that I have executed a monitored approach while flying." The responses were widely distributed: 35.1% of pilots (n=72) never executed a monitored approach; 7.8% (n=16) executed the approach one or two times; 6.3% (n=13) executed the approach three to four times; 18.5% (n=38) executed it several times; and 32.2% (n=66) executed it dozens of times. About as many pilots in the sample never executed a monitored approach as those who executed one many times. Unfortunately, the data did not reveal the source of this difference, but it is an interesting result in view of the finding that most of the pilots who completed the survey stated that they were familiar with monitored approach procedures.

It is reasonable to ask whether the pilot-monitored approach is relevant in today's aviation environment due to technological advances such as autoland capability. The pilots were asked to agree or disagree with the statement "Knowing how to execute a monitored approach may have been important in commercial aviation in the past, but it is no longer important." Although a majority of the pilots (75.6%; n=155) disagreed with the statement, 34 pilots (16.6%) were not sure and 15 pilots (7.4%) agreed (and one did not answer the question). This finding suggested that, for this sample at least, the pilot-monitored approach maintains its relevancy.

The responses indicated that the monitored approach is used to some extent and is relevant in today's aviation environment. The question then becomes, who is using it? To help identify the types of companies using the monitored approach, the survey asked pilots who had trained for and/or executed a monitored approach, to indicate the type of company for whom they worked at the time. The results suggested that a broad spectrum of organizations train pilots for the monitored approach or include it in their procedures, including:

- Major airlines
- Regional airlines
- Corporate flight departments
- Flight schools
- Cargo carriers.

Perceived Safety and Difficulty

To estimate whether pilots consider the monitored approach to be safe, the survey asked pilots to rate its safety relative to other types of precision approaches. By "other precision approaches," the survey defined any procedure a crew might follow when using precision instruments during the approach and landing. The majority of pilots indicated that it is a safe way to execute a precision approach, with 29.3% (n=60) rating it as "much more safe;" 30.7% (n=63) rated it as "somewhat more safe;" 30.7% (n=63) rated its safety about the same as a standard precision approach (7.8% (n=16) rated it as "somewhat less safe" and 1.5% (n=3) rated it as "much less safe"). Regarding its difficulty, most pilots (47.8%, n=98) rated it as "about the same" as other precision approaches. Although the second largest percentage of pilots (31.7%, n=65) rated it as easier, many pilots (19.5%, n=40) rated it as harder to execute; two respondents did not answer this question. These results should be interpreted with caution, because a pilot's opinion on this question is likely to be dependent upon his or her experience and training.

The purpose of the pilot-monitored approach is to allow each pilot to focus on his or her particular task. However, there may be difficulties inherent in this method, with some aspects being particularly difficult. Determining which aspects are more problematic for the crew is relevant to developing training protocols and procedures. To gain information about this issue, the survey asked those pilots who had executed a monitored approach to select its "most difficult" from a list of four possibilities. (Only pilots who had executed a monitored approach were asked to complete this question.) One hundred and forty-five pilots provided an answer (sixty did not), as shown in Table 2. Although a minority of respondents (37.2%;

n=54) selected “all aspects are easy” as their response, most of the respondents (60%; n=87) selected one of the possibilities as being the most difficult.

Table 2
What aspect of the monitored approach do you think is the most difficult?

Responses	Percent	Number
Visually acquiring the runway	16.1%	33
Briefing the crew before the approach	10.7%	22
Determining whether to go-around or land	9.3%	19
Following the instruments - being heads down	6.3%	13
None- all aspects are difficult	2.8%	4
None –all aspects are easy	37.2%	54

Importance of Training

The results suggested that the respondents consider training essential to the safe execution of monitored approaches. When asked to indicate what degree of discomfort they would feel if they were to execute a monitored approach with a crew that had no training on the monitored approach, 32.2% of the respondents (n=66) said they would have “high discomfort” and 42% (n=86) would have “some discomfort,” whereas 2% (n=4) would have “no discomfort” and 11.2% (n=23) would have “little discomfort.” (Some pilots (12.7% n=26) were not sure.) Several pilots commented on the need to improve training and the importance of training. One pilot stressed that “practice, practice, practice” was necessary to improve the execution of monitored approaches.

Pilots’ Comments on the Monitored Approach

The last page of the survey asked pilots to submit comments, limited to 250 words, for each of the following items:

1. Describe a difficult monitored approach that you have experienced.
2. List the strengths of the monitored approach.
3. List the weaknesses of the monitored approach.
4. In your view, how can the monitored approach method be improved?

This section provided an opportunity for pilots to submit ideas in their own words on topics that may not have been covered by the earlier survey items. One hundred and five pilots (51%) submitted comments. The comments are summarized below in Tables 3 through 8. Each table lists the comments in one column and the number of pilots making that comment is shown in the “numbers” column.

Difficult Conditions

The pilots described many conditions and/or factors that render monitored approaches unusually difficult. Difficult conditions fell into those that are either externally controlled or located, such as weather (see Table 3), or those that are related to crew management and actions (see Table 4). Other difficulties are inherent in the procedure itself (see Table 5):

Table 3
External Conditions Associated With Difficult Monitored Approaches

External Conditions	Number
Combined conditions of low visibility and strong wind	11
Low visibility	5
Windy conditions	2
Autoland not available	2
Airport location - high terrain	1

Table 4
Crew-Based Problems Associated With Difficult Monitored Approaches

Crew-Based Problems	Number
Weak/inexperienced crew member	3
Insufficient briefings	3
Crew member called for a landing before visually acquiring the runway	1
Pilot-in-Command became distracted at about 200 feet above minimums	1
Crew was slow to reconfigure aircraft for the missed approach	1
Monitoring pilot "panicked"	1
Pilot overshot the ILS	1
Pilot Monitoring distracted himself and the other pilot by calling ATC rather than on focusing on missed approach procedure	1
Pilot Monitoring hesitated to make the call for a landing or a missed approach	1
Late handover of controls	1
Handed over controls several times	1

Table 5
Difficulties Inherent in Monitored Approaches

Procedures	Number
Swapping controls at decision height	1
More concentration required	1
Division of duty	1

Strengths of the Pilot-monitored Approach

The pilots provided characteristics of the monitored approach that they consider strengths, shown in Table 6. Among the strengths cited by many respondents was its effect on crew coordination. Executing a pilot-monitored approach requires that the crew be tightly coordinated and disciplined. Similarly, executing the approach requires that the crew make explicit the roles and responsibilities of each pilot. One pilot stated that properly executing a monitored approach means that each pilot will be "completely in the loop." This sharing of responsibility also has the benefit of reducing individual workload.

Table 6
Strengths of the Monitored Approach

Grouped Comments	Number
Leads to clear and specific task assignments for each crew member	17
Facilitates the transition to visual reference/ minimizes transition time from instruments to external cues	16
Provides the ability to acquire the visual reference necessary for landing	12
Increases ability of Pilot Flying to attend to instruments (develop scan and recognize deviations)	11
Forces crew to be prepared for missed approaches	8
Forces crew to be tightly coordinated	6
Pilot with visual reference lands the airplane	6
Enhances situational awareness for the Pilot in Command	5
Provides strong back-up between crew members	5
Crew resource management (non specific)	4
Each crew member performs the task for which s/he is most experienced	3
Easy to execute	3
Lowers individual work load	3
Safer than other approaches	3
Eliminates need to switch between instruments to outside	2
Increases the ability of crew to pay attention to tasks	2
Enables the Pilot Flying to focus exclusively on the instruments	2
Enables positive control of the aircraft	1
Enhanced situational awareness	1
Forces crew to have thorough briefings	1
Minimizes the risk of the aircraft pitching up when the crew transitions from instruments to visual reference	1
No change in control at critical point	1

Weaknesses of the Pilot-monitored Approach

Table 7 summarizes pilots' comments about the weaknesses of the monitored approach. In many ways, the pilots' comments revealed that its very strengths could also be viewed as vulnerabilities. For example, the requirement that the crew be well coordinated throughout the procedure means that breakdowns in communications or inadequate briefings can have particularly negative consequences. Another perceived weakness is the changing of aircraft controls in close proximity to the ground, an aspect that was noted by several pilots as an inherent risk in the approach.

In addition to procedural and training problems, several pilots noted other, non-technical weaknesses. For example, one pilot commented that it is "sometimes difficult to put your full trust in the other pilot." Also of concern is the lack of "feel" of the aircraft by the pilot who takes control for the landing; this consideration may be especially important in conditions where there is a strong cross wind or icing, making the aircraft relatively more difficult to control. Other pilots pointed to the frustration of not flying the plane from take-off to landing, or the dissatisfaction of "not landing it yourself;" one observed that it is "not much fun." Finally, several pilots consider a lack of training and practice as a weakness.

Table 7
Weaknesses of the Monitored Approach

Potential Weaknesses	Number
Change of aircraft control at critical stage and close to the ground	22
Requires that crew be well trained	12
Requires that crew be well coordinated	8
Requires a good briefing	6
Requires that the Second-in-Command be proficient	3
The transition to visual reference is still necessary	3
Not satisfying for the Not Landing Pilot to fly, but then not land, the aircraft/ the Not Landing Pilot may not be willing to cede control to Landing Pilot	3
Short time to acquire visual reference to runway	3
Tempting to look away from the instruments to outside	2
There is no standard for training	2
Forces the Captain to watch activity of the First Officer very closely/ Restricts ability of the Captain to get additional help from the First Officer	2
It "divides the cockpit"	1
Can be distracting when company procedures require many "call outs"	1
Cannot be performed in combination with a HUD	1
Creates more workload	1
Landing Pilot cannot cross-check with course and glideslope	1
The pilot-monitored approach may not be necessary	1
Not all aircraft are properly equipped	1
Requires that companies have specific procedures	1

Improvements

Pilots recommended a diversity of improvements for the monitored approach, as shown in Table 8. Many pilots approached this question by suggesting improvements through training and standardization, whereas other pilots recommended improvements to the procedure itself. At least 19 pilots suggested that either more training or the standardization of training would be an improvement, and several pilots recommended that more practice or regular use is required. In addition, several pilots called for stronger guidelines, checklists, and procedures. Another pilot noted that standardization might be important especially for crews that have not received formal training on the monitored approach. One pilot recommended that an Advisory Circular be published. In contrast, a few pilots suggested that the monitored approach is no longer needed.

Many respondents made specific recommendations in regards to the procedure. For example, several pilots advocated that the monitored approach could be improved with the integration of Heads-up Displays (HUDs) in the cockpit. Interestingly, some pilots recommended a HUD for the Pilot Not Landing, whereas other pilots suggested that a HUD would support the Pilot Landing by improving the ability to capture the outside environment more quickly and accurately. A few pilots called for the use of an autopilot until the aircraft reaches decision height, but others flatly stated that the monitored approach should not be used with an autopilot. In fact, several pilots suggested that the monitored approach is not needed with today's flight guidance systems. In contrast, several pilots said because of the effectiveness of the monitored approach, they could not think of any ways to improve it. For example, one person wrote, "If used properly, it is an excellent technique." Another commented that she or he could not think of an improvement because, "It's brilliant."

Table 8
Pilots' Recommendations for Improving the Monitored Approach

Ways to Improve the Monitored Approach	Number
Integration with HUDs	12
Increase training requirements	7
Require that all crews be trained on it	6
More frequent use	4
Standardization	4
Standardize training	3
Inform pilots of its value in reducing crashes into terrain (CFIT)	2
Require that all/most airlines use it	2
Use of autopilot	2
Use of Enhanced Vision Systems	2
Do not use with coupled autopilots	1
Made mandatory for all non-precision approaches	1
Only have call outs for non-normal things such as "No Flare" or "No Rollout"	1
Publish strong guidelines and checklists	1
Use standardized missed approach procedures	1
Stronger crew discipline	1
Train for it using simulators	1
Use autopilot until Decision Height	1
Use the word 'Decide" at "minimums" to promote the act of either committing to land or following a missed approach procedure	1

Conclusions

The results identified several key aspects of the pilot-monitored approach. Specifically, the results showed that there is ambiguity about the meaning of the term “monitored approach” as well as uncertainty about FAA policy toward it. Nonetheless, the survey data strongly suggested that the monitored approach has a role in commercial aviation. Further, the sampled pilots in general view it as a safe and useful procedure. The results showed that aviation organizations provide some training for the monitored approach and that pilots use monitored approach procedures. Clearly, changes in technology have not made it obsolete. It remains unknown, however, what priority the monitored approach holds in the training protocols of carriers and companies. If monitored approaches are conducted very infrequently, as seems likely, then carriers and companies may not view training for it as cost-effective. However, the fact that it is executed infrequently and under difficult circumstances points to a need for training.

The results of the survey provided an initial glimpse into aspects of the monitored approach upon which further research can be built. The results revealed that there are many issues to be resolved or even identified. In general, directly asking pilots for their opinions through an Internet survey was a relatively cost-effective and fruitful endeavor that helped to address the research objectives. However, there are limitations to the research that should be kept in mind. For example, the survey could not confirm the status of respondents or randomly select the respondents. Likewise, as in other types of surveys, some respondents may doubt the security of the information that they provide, and, for this reason, may have been unwilling to share information that is sensitive or critical. In addition, although some organizations showed an eagerness to help and recruited dozens of pilots to take the survey, other organizations were not as responsive. One of the most successful avenues of contacting pilots was through internet forums, including Bluecoats, Landings, and the AirlineCrew lists, a finding which suggests that pilot forums may be a valuable resource in future research efforts. Important follow-up questions include the following:

- Should the FAA mandate the use of the pilot-monitored approach for special situations?
- Should the FAA mandate training or standardize training for the pilot-monitored approach?
- What type of training is most effective?
- What type of criterion should be applied for training of the pilot-monitored approach?
- What, if any, aspects of the monitored approach require improvement?
- Are there types of aircraft for which the pilot-monitored approach is more important? If so, on what basis can this classification be made?

To pursue these questions, one avenue would be to conduct an in-depth survey with a randomized sample that represents the nation’s population of professional pilots. A national, comprehensive survey would counter some of the limitations of the current data set and would provide details about training procedures, safety advantages, and possible improvements. It is important to identify the circumstances when the pilot-monitored approach should be used. Another fruitful endeavor would be to conduct empirical, simulator-based research in order to identify useful training protocols as well as variations in procedure that improve performance and safety in the pilot-monitored approach.

References

National Transportation Safety Board. (2000). *Controlled flight into terrain Korean Air Flight 801, Boeing 747-300, HL7468 Nimitz Hill, Guam August, 6, 1997*. Aircraft Accident Report. NTSB/ AAR-00-01. Washington, DC.

National Transportation Safety Board. (2001). *Aircraft accident brief: Ground impact of American Airlines 1340, Chicago, Illinois, February 9, 1998*. Aircraft Accident Report. NTSB/AAB-01/01. Washington, DC.

Nolan, M. S. (2004). *Fundamentals of Air Traffic Control* (4th ed.). Belmont, CA: Brooks/Cole.

International Journal of Applied Aviation Studies, 5, Number 2

Relationships among Computer-Based Instruction and Reasoning Ability on Science Students

Henry H. Baharestani
University of Oklahoma
College of Continuing Education
1700 Asp Avenue, Box 10
Norman, OK 73072

Abstract

The purpose of this study was to examine the nature and extent of the understanding of Newtonian physics among science students. This study investigated differential effects of three different instructional treatments: Computer Text Instruction (CTI), Computer Text-Graphic Instruction (CTGI), and Computer-Based Instruction (CBI). These treatments exposed areas of the subjects' lack of understanding of physics concepts, which were measured by the Force Concept Inventory (FCI) to determine the students' misunderstandings of Newtonian physics concepts. The sample consisted of 90 undergraduate students, with non-physics majors, enrolled in the "Physics for Life Science" course at the University of Oklahoma. The results indicated that students who used the CBI lesson did significantly better with respect to understanding Newton's laws than students using the CTGI or CTI lessons. Air traffic controller trainees and science students are alike in that space-time concepts and reasoning ability of facts are equal components of course curricula.

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

Purpose of the Study

Introduction

This study examined the importance of Computer-Based Instruction as a methodology of instruction for learners in diverse disciplines. Students from physics, air traffic control, and flight training have learning characteristics in common with each other (Baharestani, Strauss & Hubbard, 2001).

Numerous researchers have employed a wide variety of techniques to evaluate the overall effectiveness of computers in enhancing student classroom learning (Roblyer, Castine, & King, 1988). Interestingly, the majority of these researchers found computers quite effective for overcoming certain cognitive difficulties in the students' grasp of science concepts (Weller, 1995; Kulik & Kulik, 1991; Wise, 1989; Roblyer, Castine, & King, 1988).

In addition to instructional procedures, other factors related to student abilities and learning approaches may explain the difficulty students have with understanding physics concepts. One such factor is students' reasoning ability, ranging from pre-concrete to formal (Lawson & Thompson, 1988). Students who are capable of abstract formal reasoning can obtain a more sound understanding of abstract concepts in physics (Lawson & Thompson, 1988).

For several years, teachers have used Computer-Based Instruction to help improve their students' understanding of difficult concepts in education. This section reviews some recent research regarding the use of computer interactive strategies.

Williamson and Abraham (1995) investigated the effect of computer animation on students' visualization of chemistry concepts. Their results indicated that students who viewed the animations had fewer misconceptions of chemistry concepts. This finding suggested that Computer-Based Instruction can be effective in mastering certain kinds of conceptual understandings.

Wise (1989) conducted a meta-analysis comparing Computer-Based Instruction with traditional instructional approaches. He found that student achievement was positively affected by Computer-Based Instruction. He then divided computer usage into five categories: laboratories, tutorials, testing, simulations, and videodisc lessons. Wise (1989) found the highest effect size (ES) was with computer-based laboratories (ES = 0.76), a finding significantly higher ($p < .10$) than that of simulation and testing usage, but not different from tutorial or videodisc usage. Furthermore, the author found that the effect size was greater for the physical sciences, although not significantly different from that for biological sciences (ES = 0.45 and 0.22, respectively).

Park and Gittelman (1992) carried out a study to test the hypothesis that animated visuals are better than static visuals in enhancing the learning abilities of the learner. Their findings indicated that animated visual displays in Computer-Based Instruction were more effective than static visual displays for teaching electronic circuit troubleshooting skills.

Wainright (1989) conducted a study that compared a worksheet exercise group with a Computer-Based Instruction (CBI) group in general high school chemistry classes. The treatment group used specific microcomputer software (distributed by COMPress, Inc., for the computer activity in general chemistry), while the control group used traditional worksheets (containing exercises) for daily reinforcement activities. The author found in this study that the control group's achievement scores were significantly higher than the CBI group. He also found that the use of microcomputer materials by the experimental group did not contribute to more effective learning of concepts. In addition to instructional procedures, other factors related to student abilities and learning approaches may explain the difficulty students have with understanding physics. One such factor is students' reasoning ability. Research suggests that a student's ability to reason corresponds to the student's ability to understand concepts presented in the subject area of physics (Williams & Cavallo, 1995). This may result from the formal nature of concepts in physics that force students to use higher order reasoning abilities to build logical understanding. The study further examined students' reasoning ability and their understanding of Newtonian physics. Their research indicated that reasoning ability was correlated to students' understanding of physics concepts. Moreover, students who have low reasoning ability develop more misconceptions and poorer understandings of physics concepts. Those students who had high reasoning ability had greater understanding of physics with fewer misconceptions.

Other researchers have also found that there is a link between reasoning ability and concept understanding. Renner and Marek (1988) contrasted concrete concepts (those that can be directly experienced) with formal concepts (those that require formal reasoning ability). Other researchers have also found that formal reasoning ability is necessary to understand formal concepts (Lawson & Renner, 1975; Simpson & Marek, 1988). These results indicated a need for further investigation into the leading causes of misconceptions and the instructional approaches needed to remediate misconception among students. Furthermore, since reasoning ability is related to better understanding of physics concepts, there is a need to investigate the effects of Computer-Based Instruction with students who have lower reasoning abilities.

Many studies have been done with regard to Computer-Based Instruction, reasoning abilities. None, however, have examined all of these factors simultaneously, much less the interaction between these variables relative to students' understanding of Newtonian physics.

Purpose

The purpose of this study was to comprehend more clearly the nature and extent of students' understandings and misunderstandings of Newtonian physics. In doing so, this study investigated possible differential effects of three instructional treatments. These treatments exposed areas of students' misunderstandings of physics concepts as measured by the Force Concept Inventory (FCI), an instrument developed by Hestenes, Wells, and Swackhamer, (1992) to determine students' misunderstandings of Newtonian physics concepts. Specifically, the purposes of this study were:

(1) To determine differences in the understanding of Newtonian physics by students exposed to one of the three different computer-based instructional treatments: Computer Text Instruction (CTI), Computer Text-Graphic Instruction (CTGI), and Computer Based-Instruction (CBI).

(2) To determine and investigate the differences among reasoning ability, treatment (CTI, CTGI, and CBI), and the interaction of these variables on students' understanding of Newton's laws.

Experimental Design

Method

An experimental study was conducted using three groups: CTI, CTGI, and CBI. The control group received textual information in the CTI format. The two experimental groups received treatments in either the CTGI or CBI format, but not both. A posttest instrument, FCI, was used to assess each group.

The treatment consisted of three separate groups: (1) CTI, (2) CTGI, and (3) CBI. The first, CTI, consisted of only text. The second, CTGI, contained three parts: text, pictures related to the text, and audio. The third, CBI, contained four parts: text, pictures related to the text, animation, and audio. Only the CBI section of treatment contained animation.

All three groups (CTI, CTGI, and CBI) received instruction in a common lecture section conducted by the course professor. The subject matter of this instruction was Newton's laws.

The samples for this study consisted of students enrolled in the fall semester of the "Physics for Life Science" course at the University of Oklahoma. Overall, this course had five discussion sections ranging from 35 to 40 students with a total enrollment of 180 students. Of the 180 students, 90 students fully participated in this study.

There were three class lectures of 50 minutes per week. Newton's laws were taught in class from the second through the fourth week of the semester within the time frame of the experiment. The discussion period consisted of a 25-minute problem-solving period, followed by the experimental treatment. One hundred eighty students were randomly assigned to one of the three treatment groups (CTI, CTGI, or CBI), and exposed to twenty-five minutes of the treatment, in the fourth week of the semester in the computer lab.

During the treatment, students interacted with the critical points of Newton's laws on the computer screen. Students were not allowed to repeat the treatments. The content (e.g., text, picture, animation, and audio) of each treatment (CTI, CTGI, and CBI) was programmed by the researcher using the Authorware® Programming Language.

Descriptive statistics were generated for the responses to each of the three instruments used in this study (e.g., Test of Logical Thinking and Force Concept Inventory). To determine differences among the three treatment groups, a One-Way Analysis of Variance (ANOVA) was performed using scores on the Test of Logical Thinking (TOLT). The Force Concept Inventory was the dependent variable.

Instruments

Through the entirety of this research study, one dependent variable was measured: the level of conceptual understanding of Newton's laws. One independent variable (reasoning ability) was also measured. Two instruments were used to measure the aforementioned variables as described in the following sections:

Force Concept Inventory. The Force Concept Inventory (FCI) was developed by Hestenes, Wells, and Swackhamer (1992). It was administered immediately following the treatment in the fourth week of the semester to determine the students' misunderstandings of Newtonian physics concepts. The FCI consisted of 29 multiple-choice items designed to identify Newtonian physics misunderstandings. Questions 20 and 21 were omitted from the FCI as the students found the questions confusing in the pilot study conducted by the researcher at the University of Oklahoma. This instrument forces students to choose between correct and incorrect responses. The higher the score on the FCI, the fewer the misunderstandings and the greater the students' understanding of Newton's laws. The Kuder-Richardson reliability for the FCI is .86 if used as the pretest, and .89 if used as the posttest (Hestenes, Wells, & Swackhamer, 1992).

Test of Logical Thinking. The Test of Logical Thinking (TOLT) was used on the first day of class to determine each student's reasoning ability. The TOLT is a 10-question instrument measuring: (1) controlling variables, (2) proportional reasoning, (3) combinatorial reasoning, (4) probabilistic reasoning, and (5) correlational reasoning. Each item requires a response, along with a justification for the response. The scores on the TOLT range from 0 to 10, with 10 representing complete formal operations. A student was given one point for a correct answer and no points for an incorrect answer. A student scoring five points or less was labeled a concrete learner. A student scoring six points or higher was labeled a formal learner. Internal reliability is reported for students from grade 6 through college as .85 (Tobin & Capie, 1981). Moreover, the criterion validity between the TOLT and Piagetian interview is .80 (Tobin & Capie, 1981).

Computer Text Instruction Treatment. The Computer Text Instruction (CTI) lesson consisted only of formal text. The content of the CTI lesson included Newton's laws of motion, projectile motion, and momentum. The CTI software was kept in a general physics computer lab, and students accessed the CTI lesson following the fourth week of the discussion section. The computer lab was open to students only during the treatment.

Computer Text-Graphic Instruction Treatment. The Computer Text-Graphic Instruction (CTGI) lesson contained three parts: text, static pictures related to the text, and audio. The content of the CTGI lesson included the same content as the CTI lesson (Newton's laws of motion, projectile motion, and momentum). The CTGI lesson was also kept in a general physics computer lab, and students accessed the CTGI lesson following the fourth week of the discussion section. The computer lab was open to students only during the treatment.

Computer Based Instruction Treatment. The Computer-Based Instruction (CBI) lesson contained four parts: text, static pictures related to the text, audio, and eight different animation sequences. The content of the CBI lesson included the same content as the CTI lesson (Newton's laws of motion, projectile motion, and momentum). The CBI lesson was also kept in a general physics computer lab and students accessed the CBI lesson following the fourth week of the discussion section. The computer lab was open to students only during the treatment

The CTI, CTGI, and CBI content was reviewed by an expert panel consisting of (1) a college physics professor, (2) a high school physics teacher, (3) a science education professor, and (4) an Instructional Systems Design expert. As a result, several changes in the vocabulary of content of both Newton's laws of motion and momentum and in the animation of the CBI were made according to their recommendations. The CTI, CTGI, and CBI lessons were also pilot-tested with forty students in the summer of 1998. Of these forty students, six students from the pilot study were interviewed regarding their related interpretations and understandings of content in each treatment condition. In addition, pencil and paper were provided for students' comments regarding interpretation and understanding of content, screen graphics, and the use of animation in each treatment condition. As a result, the suggested modifications obtained from the reviews, pilot test, interviews, and pencil and paper responses were used to produce the final forms of the CTI, CTGI, and CBI lessons.

Results and Implications

Results

Question 1. To determine differences in the understanding of Newtonian physics by students exposed to one of the three different computer-based instructional treatments (CTI, CTGI, and CBI).

Concept understanding was measured by the Force Concept Inventory (FCI). The FCI was given during the fourth week of the semester following the treatment. The FCI contained 27 items with a maximum possible score of 27. Means for each of the treatment groups are in Table 1.

ANOVA results for the FCI are presented in Table 2. The ANOVA results reveal a significant difference in the FCI score by treatment. Differences at the .05 level were followed by Newman-Keuls post hoc tests to determine which pairs of the three treatment groups' means differed. The CBI group had scores significantly different from those of either the CTGI or the CTI groups. However, the score of the CTI and CTGI groups were not significantly different.

Table 1
Group Means on the FCI

Treatment	Count	Mean	Std. Deviation	Std. Error	95 Pct. Con.
CTI	30	11.73	4.18	.76	10.17 - 13.30
CTGI	30	13.96	5.54	1.01	11.90 - 16.04
CBI	30	16.90	4.22	.77	15.32 - 18.48
Total	90	14.20	5.10	.53	

Table 2
ANOVA Results for the FCI Scores of the Treatment Groups

Source	Sum of sq.	D.F. †	Mean sq.	F-value°	P ‡
Between	402.87	2	201.43	9.15	.002 *
Within	1915.53	87	22.02		
Total	2318.40	89			

* The mean difference is significant at the .05 level.
† D.F. means degrees of freedom.
‡ P is the probability that the two distributions overlap.
° F-value is the variation both within and between each of the groups analyzed statistically.

To assess the magnitude of a difference between the means of two groups is to calculate what is known as effect size (ES). The effect sizes were calculated by dividing the difference in the means of the CTI group and the CTGI or CBI group by the standard deviation of the CTI group. An effect size of .5263 was found between the CTGI group and the CBI group on the FCI. An effect size of 1.2346 was found between the CTI group and the CBI group on the FCI. The CBI treatment resulted in an increase of the mean score of about 5.1667—a standard deviation from the CTI group.

Question 2. To determine and investigate the differences among reasoning ability, treatment (CTI, CTGI, and CBI), and the interaction of these variables on students' understanding of Newton's laws.

As shown in Tables 3 and 4, no significant difference ($\underline{F}$=3.613, $\underline{P}$=.061) was found in student reasoning ability and the treatments as measured by scores of the FCI. Although not significant ($\underline{P}$=.061), the portion of the variance explained by reasoning ability was relatively high. However, Table 3 shows that reasoning ability alone accounted for 21.5% of the observed variance for the scores of the FCI. This indicated that the chance of finding significant results was likely. Thus, although the observed variance explained by the reasoning ability was relatively high, the finding of no significance gives greater relevance to these results than the percentage (21.5%) indicates.

Tables 3 and 4 show no significant difference between the means of the concrete and formal learners. The higher mean achieved by concrete learners (15.191) may indicate an overall better understanding of Newton's laws than formal learners (13.116). Those students who did not possess the reasoning ability needed to understand spatial concepts might have resorted to memorizing facts, formulas, and problem types to get through physics courses (Hammer, 1989; Hewitt, 1995; Renner & Marek, 1988). Another explanation might be that the concrete learners have resonated better with the subject matter than the formal learners.

Table 3
Univariate Analysis of Variance for the FCI Scores in the Treatment Groups vs. Concrete and Formal Reasoning Students

Source	Sum of sq.	D.F.	Mean sq.	F	P
Corrected Model	499.264*	5	99.853	4.611	.001
Intercept	17943.653	1	17943.653	828.562	.000
Groups	376.291	2	188.146	8.688	.000
Concrete/Formal	78.253	1	78.253	3.613	.061
(Groups)(Concrete)	17.671	2	8.835	.408	.666
Error	1819.136	84	21.656		
Total	20466.000	90			
Corrected Total	2318.400	89			

* R Squared = .215

Table 4
Descriptive Statistics for the FCI Scores in the Treatment Groups vs. Concrete and Formal Reasoning Students

Groups	Concrete/Formal	Mean	Std. Deviation	N
CTI	Concrete Student Formal Student Total	13.0000 10.6250 11.7333	4.8358 3.2838 4.1848	14 16 30
CTGI	Concrete Student Formal Student Total	14.2353 13.6154 13.9667	5.3564 5.9797 5.5428	17 13 30
CBI	Concrete Student Formal Student Total	18.1250 15.5000 16.9000	3.7572 4.4159 4.2210	16 14 30
Total	Concrete Student Formal Student Total	15.1915 13.1163 14.2000	5.0975 4.9435 5.1039	47 43 90

It appears from Table 3 that there is no evidence of interaction ($\underline{P}$<.666) between the treatment and reasoning ability variables on students' understanding of Newton's laws.

Discussion and Implication for Further Research

Question 1 Discussion

Based on effect size a difference in ability to respond to test items did exist between CTI, CTGI, and CBI groups. The nature of the instructional design and method appeared to contribute more to the results than any other factor accounted for in this study. Perhaps a more defined distribution would have occurred if the learners were assessed according to learning style or type of intelligence within each treatment group. It would appear that students not possessing an enhanced spatial intelligence found two-dimensional drawings (CTGI) incomplete. The CTI format would limit the non-spatial student perhaps to an even greater degree. However, the CBI format provided the spatial and non-spatial students with all they needed to succeed. An animation of the actual teaching point was clearly displayed and then tested. These results are generally consistent with other studies reported in literature (Aiello & Wolfle, 1980; Kulik & Kulik, 1991; Roblyer, Castine, & King, 1988; Wise, 1989; Gardner, 1993).

Computer-Based Instruction, if designed correctly, incorporates features that attract all learning styles. Animations or video clips are certainly the most dynamic teaching tools but are only pieces of the cognitive puzzle. CBI uses the strength of textual information, which in this study, existed exclusively in the CTI format to probe the memory and experiences of each student. The students cognitively interacted with the information on the screen by combining past perceptions of the world around them with the current information on Newton's laws introduced by the CBI treatment. Simple graphics, a design feature used in this study's CTGI format, enhanced the CTGI and CBI presentations by giving the students the same frame of reference. However, where CBI pulls ahead of CTI and CTGI is in its ability to combine text, graphics, and animation during the instructional phase to more completely explain physics concepts.

Question 2 Discussion

There was not a significant difference among reasoning ability of students and the treatments as measured by the scores of the FCI. This finding contradicts the results of others that suggested that students who had high reasoning ability had the greatest physics understanding and fewer misconceptions (Williams & Cavallo, 1995).

There was evidence of pre-learning in the sample groups. The students exhibited a baseline rational ability (R=21.5%). The treatments built on that baseline ability in one of three ways: text only (CTI); text and graphics (CTGI); and text, graphics, and animation (CBI).

The affect of teacher-student interaction during classroom instruction was not accounted for in this study. This should not have affected the results, in that all of the students received the same amount of classroom instruction. What is purely individual is one's motivation to learn, natural cognitive ability, and life experiences. In addition, mixing qualitative research with the quantitative was not done and should be investigated in another study.

Effect sizes distinguished differences between treatments more than any other statistical analysis. Rationality differences were not significantly different between concrete and formal groups, offering evidence that the groups were rationally homogeneous. The real difference was clearly in instructional design, with animation within the CBI format showing the greatest advantage when teaching physics concepts.

The logical tie between physics subjects and air traffic controllers lies in the commonality among reasoning ability in learners across areas of interest. By focusing on reasoning ability, the air traffic controller might be better equipped to handle unusual problems not previously practiced in the laboratory. Moreover, Computer-Based Instruction provides an opportunity for interaction and excites visual, auditory, and tactile sensory responses. Since CBI incorporates multimodal cueing, it would appear that this form of instruction has a greater resident capability to improve students' understanding of abstract concepts in physics (Baharestani, Strauss & Hubbard, 2001). Air traffic controller trainees emerge to have the same learning characteristics as subjects in the Newtonian physics study.

Conclusions

Physics knowledge, in most cases, depends on the student's ability to imagine operations in space-time without necessarily seeing those operations. Results of this study indicated that students who are capable of abstract formal reasoning could obtain a more sound understanding of abstract concepts in physics.

Most of the literature on teaching science stresses experiments and demonstrations that are very helpful to students who have a willingness to learn science concepts. However, there is less debate about modeling the content and delivery of science curricula to become more attractive to students who lack an interest in science concepts (Knupfer & Zollman, 1994).

The conceptual understanding of Newton's laws as measured by the Force Concept Inventory (FCI) scores was significantly increased for students who interacted with Computer-Based Instruction with animation within the CBI format. Effect sizes of approximately 1.2 were found. It is evident that the use of animation in teaching Newtonian physics can improve a students' overall conceptual understanding of physics. The computerized visual animation in the CBI helped physics students develop better understanding of Newtonian physics concepts. These findings are consistent with the current literature that studies how computerized visual animation makes concepts more accessible to science students (Escalada & Zollman, 1997).

Air traffic controllers and pilots are similar to the researcher's group of science students in that all three groups must imaginatively perceive concepts in space-time while also thinking and recalling facts (Baharestani, 1999; Baharestani, Strauss & Hubbard, 2001).

Finally, the results of this study suggested that the use of animation in Computer-Based Instruction embedded in the curricula can be a very important part of teaching. Further research is needed to confirm these findings with different samples, such as a larger number of students in air traffic controller and flight training.

References

Aiello, N.C., & Wolfle, L.M. (1980, April). *A meta-analysis of individualized instruction in science.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 190 404).

Baharestani, H. (1999). Relationship among reasoning ability, meaningful learning, and computer-based instruction on students' understanding of Newton's laws. *Dissertation Abstracts International, 99,* 35529, (University Microfilms No. AAG-9935529) University of Oklahoma, Norman, OK.

Baharestani, H.H., Strauss, M.G., & Hubbard, T.P. (2001). Computer-based instruction: A common medium for divers groups. *International Journal of Aviation Research and Development, 1*(1), 19-28.

Escalada, L. T., & Zollman, D. A. (1997). An investigation on the effect of using interactive digital video in a physics classroom on student learning and attitudes. *Journal of Research in Science Teaching, 34*(5), 467-489.

Gardner, H., (1993). *Frames of mind: The theory of multiple intelligences.* New York, NY: BasicBooks.

Hammer, D. (1989). Two approaches to learning physics. *The Physics Teacher, 27*(9), 664-670.

Hestenes, & Wells, & Swackhamer (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141-158.

Hewitt, P. G. (1995). Lessons from Lily on the introductory course. *Physics Today, 48*(9), 85-86.

Knupfer, N. N., & Zollman, D. (1994) *Using DVI to teach physics: making the abstract more concrete.* Paper presented at the annual meeting of the National Convention of the Association for Educational Communications and Technology, Nashville, TN.

Kulik, C-L. C., & Kulik, J.A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*(1-2), 75-94.

Lawson, A.E., & Thompson, L.D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of Research in Science Teaching, 25*(9), 733-746.

Lawson, A.E., & Renner, J.W. (1975). Relationships of science subject matter and developmental levels of learners. *Journal of Research in Science Teaching, 12*(4), 347-358.

Park, O. C., & Gittelman, S. S. (1992). Selective use of animation and feedback in computer-based instruction. *Educational Technology Research and Development, 40*(4), 27-38.

Renner, J.W., & Marek, E.A. (1988). *The learning cycle and elementary school science teaching.* Portsmouth, NH: Heinemann.

Roblyer, M.D., Castine, W.H., & King, F.J. (1988). Assessing the impact of computer-based instruction: A review of recent research. *Computers in the Schools, 5*(3/4), 11-149.

Simpson, W.D., & Marek, E.A. (1988). Understanding and misconceptions of biology concepts held by students attending small high schools and students attending large high schools. *Journal of Research in Science Teaching, 25*(5), 361-374.

Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. *Educational and Psychological Measurement, 41*(2), 413-423.

Wainright, C. L. (1989). The effectiveness of a computer-assisted instruction package in high school chemistry. *Journal of Research in Science Teaching, 26*(4), 275-290.

Weller, H. G., (1995). Diagnosing and altering three Aristotelian alternative conceptions in dynamics: Microcomputer simulations of scientific models. *Journal of Research in Science Teaching, 32*(3), 271-290.

Williams, K.A. & Cavallo, A.M.L. (1995). Relationships between reasoning ability, meaningful learning, and students' understanding of physics concepts. *Journal of College Science Teaching, 24*(5), 311-314.

Williamson, V. M., & Abraham, M. R., (1995). The effects of computer animation on the particulate mental models of college chemistry students. *Journal of Research in Science Teaching, 32*(5), 521-534.

Wise, K. C. (1989). The effect of using computing technologies in science instruction: A synthesis of classroom-based research. In J.D. Ellis (ED), *1988 AETS Yearbook: Information Technology and Science Education* (pp. 105-118). Association for the Education of Teachers in Science.

International Journal of Applied Aviation Studies, 5, Number 2

Pilots' Workload, Situation Awareness, and Trust During Weather Events as a Function of Time Pressure, Role Assignment, Pilots' Rank, Weather Display, and Weather System

Ernesto A. Bustamante, Corey K. Fallon, James P. Bliss,
William R. Bailey III and Brittany L. Anderson
Old Dominion University
Norfolk, VA

Address all correspondence:
Ernesto A. Bustamante
Old Dominion University
Department of Psychology
Norfolk, VA 23529-0267
ebustama@odu.edu
(757) 675-0289

Abstract

Despite advances in sensor technology and information processing algorithms, weather forecasting remains unreliable. The reliability, consistency, and dependability of weather systems play an essential role in pilots' decision making under critical conditions. The goal of this study was to examine pilots' workload, situation awareness (SA), and trust in weather systems during critical weather events as a function of time pressure, role assignment, pilots' rank, weather display, and weather system. Results partially supported our hypotheses. Pilots' workload significantly increased as they approached the weather event. Consistent with previous research, Captains reported lower SA than First Officers (FO). As expected, when the NEXRAD system failed to provide an indication of the weather event at the specified waypoint, pilots' SA decreased as they approached the weather threat. As predicted, pilots trusted the onboard system more than the NEXRAD system, particularly when these systems displayed conflicting information as pilots' approached the weather threat. Our findings have important implications for the field of commercial aviation. Airlines should consider a change in role assignment philosophy. Airlines should also consider encouraging pilots to make deviation decisions around weather events as soon as they notice them. Last, our findings showed support for the added benefit of providing pilots with broader information regarding the potential weather threat using the NEXRAD system. Future research efforts should explore improvements to data link technology so that NEXRAD

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

information can be presented in real time. Such technological improvements may increase the reliability, believability, and dependability of the NEXRAD system, and ultimately avoid the distrust associated with conflicting information.

Pilots' Workload, Situation Awareness, and Trust During Weather Events as a Function of Time Pressure, Role Assignment, Pilots' Rank, Weather Display, and Weather System

Despite advances in sensor technology and information processing algorithms, weather forecasting remains unreliable. As a result, weather presentation displays traditionally have lacked credibility (Lindholm, 1999). The reliability, consistency, and dependability of weather displays play an essential role in pilots' decision making under critical conditions. Although environmental cues, such as cloud formation, provide pilots with information about potential weather problems, pilots must still rely heavily on information presented on cockpit displays (Stokes & Wickens, 1988). The ultimate goal of display designers is to provide pilots with the information necessary to make accurate decisions. High situation awareness (SA), a manageable level of workload, and trust in the weather displays often can facilitate accurate decisions.

The goal of this study was to examine pilots' workload, SA, and trust in weather displays during critical weather events. Pilots' workload, SA, and trust could vary as a function of several factors. In this study, we focused on time pressure, role assignment, pilots' rank, weather display, and weather system.

Workload

Mental resources are limited to the amount of information humans can process at any given time (Lysaght et al., 1989). Workload is defined as the load of information placed on these processes (Meshkati, Hancock & Rahimi, 1992). If workload becomes too great, the brain ceases to process information at the desired rate, and task performance degrades. For example, several studies indicated that high workload conditions significantly increase human errors in the cockpit (Hart & Bortolussi, 1984; Wiener & Nagel, 1988). Performance degradation is especially problematic in aviation, where there is little room for error and lives are usually at stake.

One technique for measuring workload is the subjective rating scale. This method involves asking participants to rate the amount of workload required for them to accomplish a task or a set of tasks. Subjective rating scales are easy to administer and have high face validity (Weirwille & Eggemeier, 1993). In addition, subjective workload measures are often reliable and share significant convergent validity with performance measures (Tsang & Wilson, 1997). One commonly implemented subjective workload measure is the NASA Task Load Index (TLX; Hart & Staveland, 1988).

Situation Awareness

Endsley (2000) defined SA as the perception of elements in a particular time and space as well as an understanding of their meaning and projection of their status in the future. SA has been conceptualized as having three levels: perception, comprehension and projection (Endsley, 2000). Perception is the extent to which a person can understand critical elements in the environment. Comprehension is the ability to integrate these elements with the pilot's goals. The highest level of SA is projection. This level involves the ability to predict what will happen to these elements in the near future (Endsley, 1999).

The environment's critical elements also can be subdivided. In the context of aviation, these elements fall into several classes. Some elements are geographical, including awareness of waypoints, terrain, airports, and other aircraft. Elements also can be spatial and temporal, including attitude, altitude, heading, flight path, and deviation from the flight plan. System elements are another class of element and include awareness of system status and function, settings on the radio, altimeter, and transponder, and awareness of air traffic control communication. A fourth class includes elements that exist in the environment. These elements consist of current and projected weather conditions, instrument or visual flight rules, and areas and altitudes to avoid. A final class includes tactical elements, such as identification, status, location and flight dynamics of the other aircraft, and threat prioritizing (Endsley, 1999). It is essential to consider pilots' SA for the design and evaluation of aircraft systems because loss of SA is a significant contributor to pilot error (Jones & Endsley, 1996).

A standard method for measuring SA is the Situation Awareness Rating Technique (SART; Taylor, 1989). The SART is a subjective measure that was developed based on input from experienced pilots. It includes ten items that can be clustered into three dimensions, including attentional demand, attentional supply, and understanding. The attentional demand cluster includes instability, variability, and complexity of the situation. Attentional supply includes arousal, spare mental capacity, concentration, and division of attention. The third dimension, understanding, includes information quantity, quality, and familiarity (Jones, 2000).

Trust

Trust plays a crucial role in the effective interaction between humans and automation. Trust helps pilots cope with the complexity of automated systems by increasing their reliance and compliance (Lee & See, 2004). By relying on the automation and complying with its directives, pilots can reduce the workload associated with supervising system behavior. This reduction in workload is particularly important when the task is complex, uncertain, or time-constrained. When the task is complex, it is impractical for pilots to allocate a large proportion of their attentional resources to monitoring automated systems. Similarly, when the task is uncertain, or pilots are under time pressure, they may have difficulty analyzing all relevant information before making decisions. By increasing compliance and reliance, trust can save pilots time and resources and ultimately improve their decision-making accuracy. However, as Lee and See (2004) emphasized, it is important to design systems for adequate trust. When systems are unreliable, high levels of trust may actually hinder human performance and compromise safety.

Time Pressure

The aircraft's distance from the weather event may also affect pilots' workload, SA, and trust given its direct relationship with time pressure. As the aircraft approaches a potential weather threat, pilots have less time to respond and make a safe deviation. This increase in time pressure may cause a subsequent increase in pilots' workload.

In addition, as the aircraft travels closer to the potential weather threat, the weather begins to enter the onboard system's range. The newly available weather information may result in increased pilot SA. On the other hand, the increase in time pressure may actually reduce SA. Decision making researchers have found that time pressure can alter the decision making process.

As the aircraft approaches the potential weather threat, the information presented on the onboard system may be more accurate than the information on the NEXRAD display. This, in turn, may create two problems. First, differences in information reliability may cause the two systems to present conflicting information. Second, although pilots may be able to rely more heavily on the onboard system to make tactical deviations as they approach the weather threat, they may not be able to use the information presented by the NEXRAD system to make strategic deviation decisions.

Role Assignment and Pilots' Rank

There are two main roles that air transport pilots may perform during a flight (Jentsch, Barnett, Bowers, & Salas, 1999). One of those roles is to fly the aircraft by manipulating the primary flight controls. This role is commonly referred to as the pilot-flying (PF) role. The PF is responsible for the movement of the aircraft and maintaining the course of the flight through making heading, altitude, and airspeed adjustments, and changes to waypoints. The second role involves performing functions other than controlling the movement of the aircraft, such as communicating with air traffic control, navigating, planning the flight, and monitoring cockpit instruments. This role is commonly known as the pilot-not-flying (PNF) role. However, it is important to note that in highly automated cockpits, although the PF's role may primarily consist of monitoring the autopilot, he or she is still considered the PF even when the autopilot is engaged. The main difference between the two roles is that the PF is responsible for the movement of the aircraft, whereas the PNF is responsible for assisting the PF.

Role assignment plays an important function in determining pilots' workload and SA. Role assignment may ultimately determine the type and amount of information to which a pilot may have access at any given time. Because the PNF is not responsible for the movement of the aircraft and primarily performs a monitoring role, he or she may have direct access to more information than the PF. As a result, role assignment may determine who is better equipped to make strategic and tactical decisions, particularly during emergency situations.

It is important to differentiate between a pilot's rank and his or her role assignment because they are two completely independent concepts. In commercial aviation, a pilots' rank refers to whether he or she is a Captain or a First Officer (FO). Although most task responsibilities vary according to role assignment, the Captain is legally responsible for the flight. Some researchers suggested that SA is easier to maintain when pilots are directly in control of the aircraft (Endsley & Rogers, 1996; Sarter & Woods, 1995). However, the added workload associated with operating the flight controls may take away attentional resources necessary to maintain SA. Consequently, Captains have a higher tendency to lose SA when they are flying the airplane (Jentsch et al., 1999).

Weather Display and Weather System

Pilots must often make deviation decisions to avoid potential weather threats based on several different sources of information. Most weather displays present information solely through text. Pilots are trained to interpret textual information and integrate it with other sources of information, such as the known ceiling and visibility minima at the destination, the kind of ice that the aircraft can withstand, or the severity of turbulence. Therefore, pilots do not necessarily need visual depictions of weather threats to make decisions. However, textual presentation formats require pilots to first interpret the information, and then integrate the sources to form a mental model of the situation. By allocating the tasks of information interpretation and integration to the pilot, text displays may increase pilot workload and reduce SA.

Research suggests that information presented in an integrated format is more effective than information presented through different sources (O'Brien & Wickens, 1997). Nevertheless, designers should be cautious when presenting integrated visual information because this format may clutter the display space and hinder pilots' ability to interpret information (Lindholm, 1999; Wickens, Kroft, & Yeh, 2000).

Sophisticated displays, such as Next Generation Radar (NEXRAD), integrate weather information and present it to pilots in a graphical format. NEXRAD uses Doppler technology to assess and present wind and precipitation information on a single graphical display. This type of display technology reduces the workload associated with integrating text and ultimately enhances pilots' SA (Wickens, 2000). In addition, NEXRAD allows pilots to determine weather trends and make predictions about future developments (Boyer, Campbell, May, Merwin, & Wickens, 1995). However, the display does not present weather information in real time. NEXRAD displays update weather information approximately every five minutes (Isaminger & Proseus, 2000). Because NEXRAD is unable to keep pace with rapidly changing weather conditions, the weather information displayed to the pilots is outdated. This outdated information may be inaccurate, and therefore unreliable. The extent of NEXRAD's unreliability is partly dependent on the amount of time since its last update (Sherman, 2003). Research suggests that unreliable systems such as NEXRAD may degrade pilot trust and decision-making accuracy (Bliss, Jeans & Prioux, 1996; Gupta, Bisantz & Singh, 2002).

In contrast to NEXRAD systems, there is a weather system that relies on sensors fixed to the aircraft. This type of onboard weather system presents information that is visually integrated in real time. Therefore, the information provided by this system tends to be more reliable. However, its major drawback is the scope of the information provided. Unlike NEXRAD, the onboard system displays a limited array of weather information directly in front of the aircraft. Pilots are typically aware of this system's limitations and do not misuse the automation. However, the onboard system's limitations may restrict pilots' ability to make more accurate tactical and strategic decisions about upcoming weather threats.

The NEXRAD and onboard systems are designed to complement each other by providing pilots with various sources of weather information. However, given the different levels of reliability associated with the NEXRAD and onboard systems, disagreement between these two sources can occur. The information NEXRAD presents may not agree with the information presented by the onboard system, particularly when the NEXRAD display has not been updated for a long time. One of the primary functions of advanced cockpit displays is to enhance pilots' SA. However, this goal is difficult to achieve when pilots are presented with conflicting information (Pritchett, 1998). Disagreement between these two weather sources may increase workload, reduce SA, decrease system trust, and ultimately degrade decision-making accuracy.

As noted above, variations in workload, SA, and trust are possible when pilots integrate information from onboard and NEXRAD displays concurrently. However, little empirical work has been done to confirm this. We conducted the current research to address these questions. It is important to note that in real-life operations, pilots' workload, SA, and trust during weather events may be affected by a number of other factors not considered in this study. Such factors may include the initial weather briefing that the FAA requires for all IFR flights and en route information provided by air traffic controllers, flight service specialists, and the airline's dispatcher. However, in this study, we were only interested in the effects of time pressure, role assignment, pilots' rank, weather display, and weather system in isolation of such other factors. In addition, it is important to emphasize that in this study, we examined pilots' workload, SA, and trust only during en route flight operations.

Hypotheses

Workload. We expected pilots to experience higher workload as a function of time pressure, role assignment, and weather display. We hypothesized that pilots would experience higher workload as they approached the weather threat (Sly & Harmann, 1999), particularly when the weather displays presented conflicting information (Pritchett, 1998), and when pilots assumed the PF role (Jentsch et al., 1999).

Situation Awareness. We expected pilots to experience lower SA regarding the weather threat as a function of time pressure, role assignment, pilots' rank, and weather display. We hypothesized that Captains would experience lower SA when they were taking on the PF role (Jentsch et al., 1999), especially when the weather displays presented conflicting information (Pritchett, 1998). Although we expected a significant effect of time pressure on SA, we made no prediction about the specific direction of the effect.

Trust. We expected pilots to trust the onboard system more than the NEXRAD system as they approached the weather event (Gupta et al., 2002), particularly when they displayed conflicting information (Pritchett, 1998).

Method

Experimental Design

We used a 4 x 2 x 2 x 3 mixed design to examine the effects of time pressure, role assignment (PF, PNF), pilots' rank (Captain, FO), and weather display (Both, Onboard, NEXRAD) on pilots' workload and SA. We manipulated time pressure by presenting pilots with information on the two weather displays at four different distances from the potential weather threat (20 nm, 40 nm, 80 nm, 160 nm). We manipulated role assignment, weather display, and weather system within groups, and pilots' rank between groups. We fully counterbalanced role assignment with pilots' rank to avoid confounding effects. We used a similar experimental design to examine pilots' trust. However, we included the type of weather system (Onboard, NEXRAD) as an additional within-groups independent variable.

Participants

We collected data from 24 pilots representing six airlines, though the majority came from United Airlines (see Figure 1). All of the pilots were male. Twelve of them were Captains and 12 were FOs. We randomly assigned them to teams (flight crews) consisting of a Captain and a FO. Captains' age ranged from 46 to 60 years (M = 55.13, SD = 4.21), whereas FOs' age ranged from 34 to 56 years (M = 46.33, SD = 5.79). The number of reported hours of glass cockpit experience ranged from 1,100 to 12,000 hours, and the number of pilot flight hours ranged from 5,000 to 19,000.

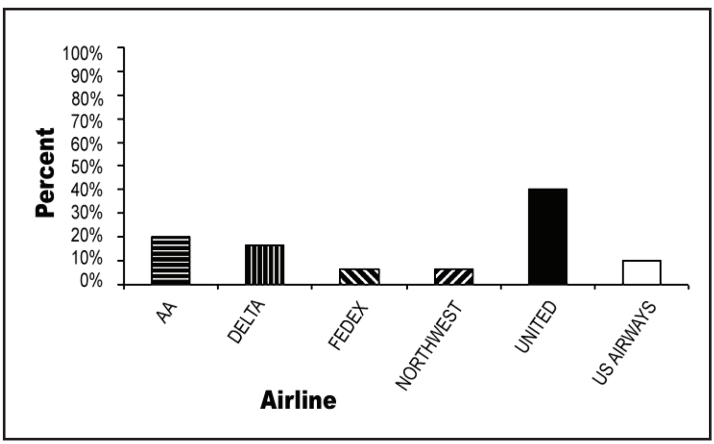


Figure 1. Pilots as a function of airline.

Materials

We used three computer workstations for this study. One computer hosted Microsoft Flight Simulator 2004 and was equipped with Flight Link's fixed wing hardware. For all flights, flight dynamics within Microsoft Flight Simulator were modeled after a Boeing 737 aircraft. A second computer to the right of the flight simulator hosted a Visual Basic 6.0 program that displayed two sources of weather information throughout the course of the flight. One source of weather information was a static image of the onboard weather radar (See Figure 2).

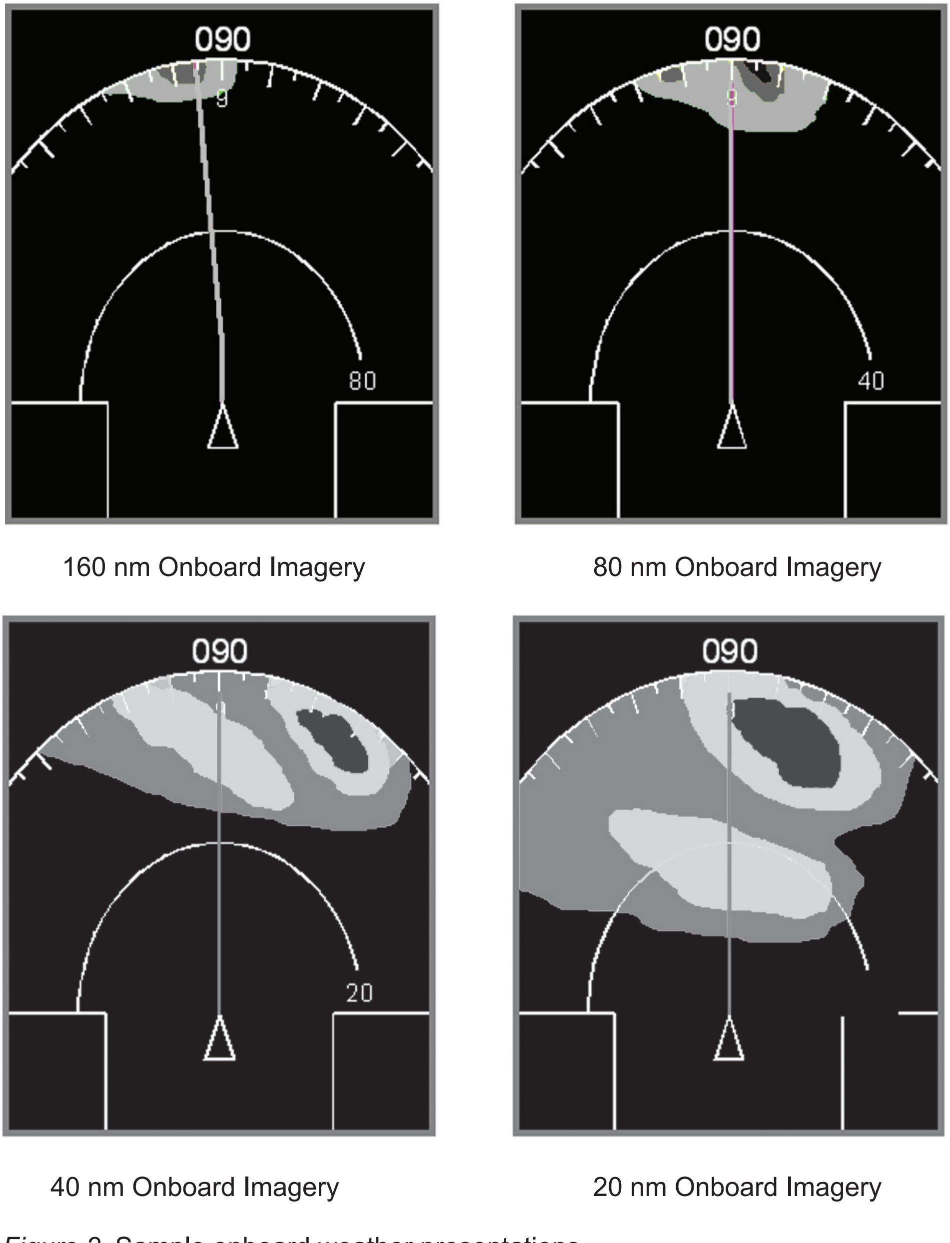
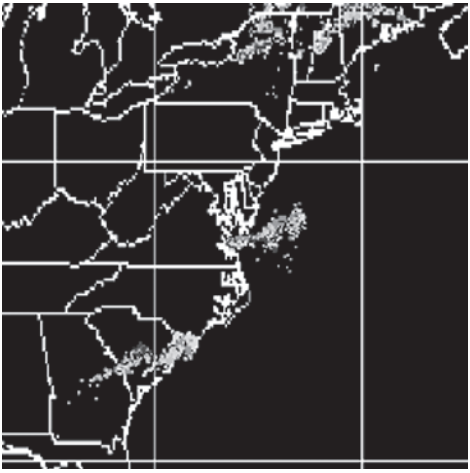


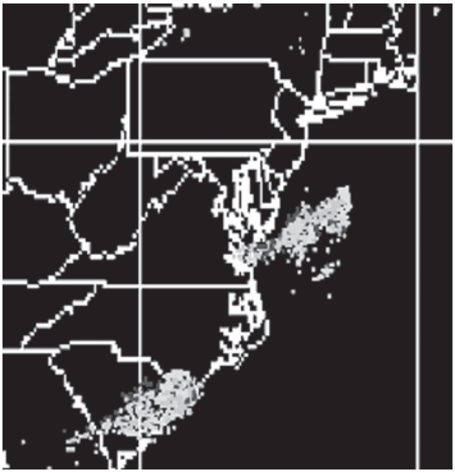
Figure 2. Sample onboard weather presentations.

The other source was a static image of NEXRAD weather imagery (see Figure 3). The NEXRAD imagery was obtained from the National Environmental Satellite, Data, and Information Service. This computer also hosted a background questionnaire, a series of deviation questions, the SART, the NASA TLX, and a 10-item trust questionnaire created by the researchers.

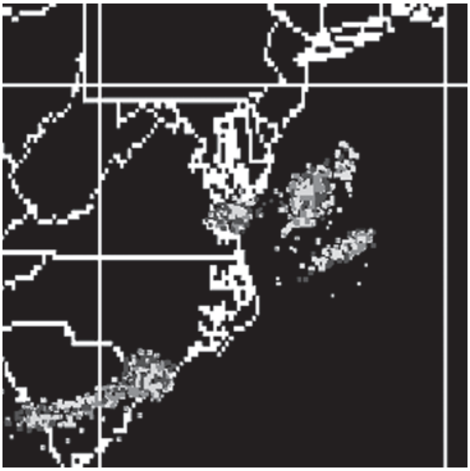
160 nm NEXRAD Imagery



80 nm NEXRAD Imagery



40 nm NEXRAD Imagery



20 nm NEXRAD Imagery

Figure 3. Sample NEXRAD weather presentations.

A third computer was positioned 90º to the left of the flight simulator. This computer also hosted the trust, workload, and situation awareness questionnaires. All computers featured Intel Pentium IV processors and 17-inch flat-panel screen monitors. Pilots completed all computerized questionnaires using a standard QWERTY keyboard and mouse.

Prior to each flight leg, pilots also received preflight briefing information. This information included the flight path and a minimal packet of weather information. The weather packet included information such as wind speed, direction, and convective activity along the projected flight path. Experimenters informed pilots that this information was 8 hours old. The usefulness of this information was limited by its age to ensure that pilots would focus more on the weather displays.

Dependent Measures

Workload and SA. We used the NASA TLX (Hart & Staveland, 1988) and the SART (Taylor, 1989) to measure workload and SA, respectively. Research suggests that subjective measures serve as effective methods for assessing workload and SA because of their close relationship with performance measures (Borg, 1978). Despite this close relationship, researchers also suggest that there are dissociations between performance and subjective measures (Endsley, 1995; Yeh & Wickens, 1988). Performance measures are sometimes insensitive to workload changes because pilots strive to maintain optimal performance even as workload increases by investing more resources into the task.

Trust. We designed a 10-item scale to assess pilots' trust in the two sources of weather information. Pilots used a continuous scale ranging from 0 to 100 to rate the degree to which they believed the onboard or NEXRAD system was inconsistent, unpredictable, truthful, accurate, trustworthy, misleading, deceptive, credible, valid, and dependable.

Procedure

The study required approximately 7.5 hours of participation from each flight crew. When the pilots arrived, they read and signed an informed consent form. Next, experimenters administered a background questionnaire and randomly assigned pilots to one of the two roles (i.e., PF, PNF). The pilot assigned to the PF role sat in front of the computer hosting Microsoft Flight Simulator, and the pilot assigned to the PNF role sat in front of the computer that displayed the weather information.

Experimenters provided the team with a brief overview of the study and administered written instructions to help pilots properly complete the NASA TLX, SART, and trust questionnaires. Pilots read the instructions and could refer back to them throughout the study.

To familiarize the pilots and reduce practice effects, experimenters instructed them to fly a practice flight leg first from Sacramento, CA to Los Angeles, CA. Before the flight, experimenters administered the preflight briefing information. The pilots were not required to takeoff or land the aircraft. They were merely required to sustain cruise flight along the flight path. During the practice flight, the teams were instructed to maintain an altitude of 19,000 ft and an airspeed of 325 knots.

During most of the flight, the weather display computer did not present any information on the monitor. The program displayed weather information only at set distances from potential weather events. Weather events represented potential thunderstorms at specific waypoints that were considered threats to flight safety. When the aircraft was 160 nm miles away from a potential weather threat, the weather display computer presented both types of weather systems. The onboard system presented weather information from the pilots' point of view, and it was presented again as the aircraft approached the weather threat at 80, 40, and 20 nm from the weather threat. The NEXRAD system presented weather information from a "god's eye" point of view, and it was presented again at 80, 40, and 20 nm from the weather threat. The NEXRAD system updated information as it approached the weather threat by zooming in the specific waypoint, thereby providing pilots with more resolution of the area. Although both systems presented static images, experimenters informed pilots that the onboard system was presenting information in real time, whereas the NEXRAD system was presenting information that might not be updated.

In the practice session, pilots encountered one potential weather threat. First, at 160 nm NEXRAD and onboard information was presented on the weather display monitor. At this point, the PF was instructed to disengage the autopilot and fly the aircraft manually to simulate the increased workload that a PF would experience during an actual weather threat. The Captain and FO worked as a team to complete a series of deviation questions based on the two sources of information. Although the pilots were instructed to work together, experimenters reminded them that the Captain must give final approval of any decision. Pilots were allotted 3.5 minutes to answer four deviation questions. Three of the questions required pilots to rate their confidence that a weather threat actually existed, their confidence that they should deviate, and their confidence in their decision on a 0 to 100 continuous rating scale. The other question required pilots to decide whether they should deviate to the right, left, or not deviate at all. These questions were geared toward addressing pilots' decision-making performance. We used these questions instead of allowing pilots to deviate from the flight path to maintain experimental control. We analyzed the results of these data and reported them at the 49th Annual Meeting of the Human Factors and Ergonomics Society (Bailey, Fallon, Bliss, & Bustamante, 2005).

After completing the deviation questions, the PF was instructed to pause the flight simulator, and both team members completed the NASA TLX, SART, and trust questionnaires independently on separate computers. The PF completed his questionnaire on the computer located 90º to his left, and the PNF completed these measures on the weather display computer. The pilots completed the NASA TLX, SART, and two trust questionnaires, one for each source of weather information.

Once the pilots completed their computerized questionnaires, the PF took his position at the flight simulator and resumed the flight. The team reengaged the autopilot and continued along the flight path. The pilots were not permitted to actually deviate from the flight path. As the team approached the weather event, they received three more presentations of the weather at 80, 40, and 20 nm from the event. The pilots followed the same procedure for every presentation.

After the practice flight leg, the pilots took a short break and experimenters answered any questions before beginning the experimental flight legs. The experimental procedure was identical to the practice session except that the pilots encountered three weather events per experimental flight leg for each level of the weather display independent variable (Both, Onboard, NEXRAD).

When both systems indicated a potential weather event at the same waypoint, this corresponded to the level of weather display referred to as "both." When only the onboard system indicated a potential weather event at the specified waypoint, this corresponded to the level of weather display referred to as "onboard." When only the NEXRAD system indicated a potential weather event at the specified waypoint, this corresponded to the level of weather display referred to as "NEXRAD." It is important to clarify that the weather display manipulation referred to which system indicated the presence of a weather threat at the specified waypoint. The weather system manipulation referred to which system (i.e., Onboard, NEXRAD) was the source of information. However, both weather systems were presented at all conditions regardless of whether or not they displayed a weather threat at the specified waypoint.

The first experimental flight leg was a flight from New York, NY to Miami, FL. The second experimental flight leg was a flight back to New York, NY from Miami, FL. After completion of the first flight leg, the pilots took a one-hr break for lunch and then reconvened for the second experimental flight leg. The Captain and FO switched roles for the second flight leg. Once pilots completed both experimental flights, experimenters debriefed and dismissed them.

Results

We first examined the effects of time pressure, role assignment, pilots' rank, and weather display on workload and SA using a 4 x 2 x 2 x 3 mixed MANOVA. Results showed statistically significant multivariate main effects of time pressure, $F(6, 17) = 4.48$, $p < .01$, Wilks' $\lambda = .39$, partial $\eta^2 = .61$, power = .92, and weather display, $F(4, 19) = 11.14$, $p < .001$, Wilks' $\lambda = .30$, partial $\eta^2 = .70$, power = 1.00.

Workload

We examined the effects of time pressure, role assignment, pilots' rank, and weather display on workload using a 4 x 2 x 2 x 3 mixed ANOVA. Results showed a statistically significant main effect of time pressure, $F(3, 66) = 8.33$, $p < .001$, partial $\eta^2 = .28$, power = .99. Follow-up pairwise comparisons showed that pilots' workload significantly increased as a function of distance from 160 nm ($M = 26.45$, $SD = 17.98$) to 20 nm ($M = 29.76$, $SD = 18.18$). These results are graphically depicted in Figure 4.

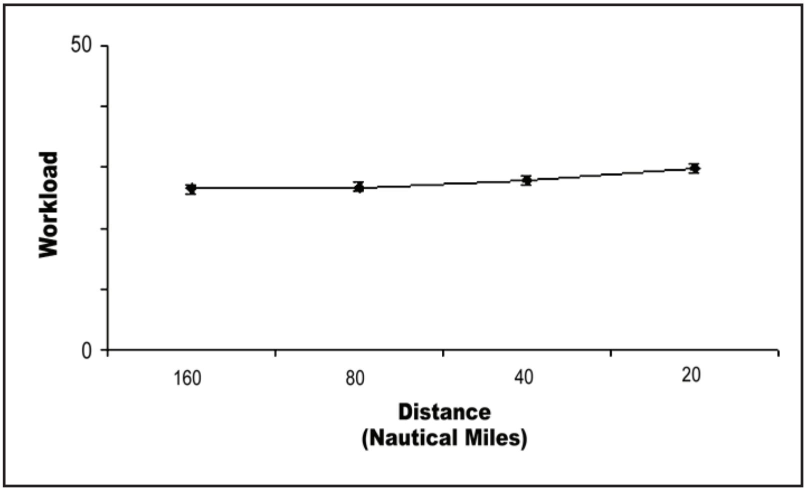


Figure 4. Pilots' workload as a function of distance from the weather event.

Situation Awareness

We examined the effects of time pressure, role assignment, pilots' rank, and weather display on SA using a 4 x 2 x 2 x 3 mixed ANOVA. Results showed a statistically significant two-way interaction effect between time pressure and weather display, $F(6, 132) = 2.81$, $p < .05$, partial $\eta^2 = .11$, power = .87. Simple effect follow-ups showed that when the NEXRAD system did not provide pilots with an indication of an upcoming weather event at the specified waypoint, pilots' SA significantly decreased as they grew closer to the potential weather event, $F(3, 141) = 7.56$, $p < .01$, partial $\eta^2 = .14$, power = .99. These results are graphically depicted in Figure 5.

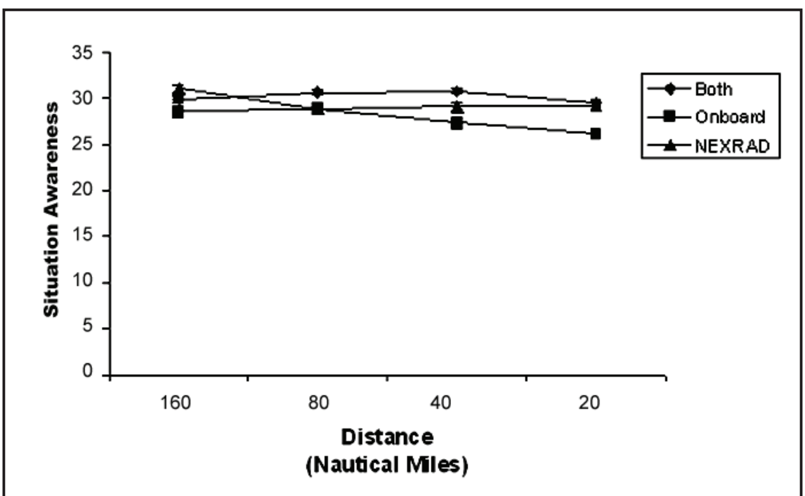


Figure 5. Situation awareness as a function of distance from the weather event and weather display.

Last, results showed a statistically significant main effect of pilots' rank, $F(1, 22) = 5.34$, $p < .05$, partial $\eta^2 = .20$, power = .60. FOs reported a significantly higher level of SA ($M = 32.10$, $SD = 7.73$) than Captains ($M = 26.15$, $SD = 7.20$). These results are graphically depicted in Figure 6.

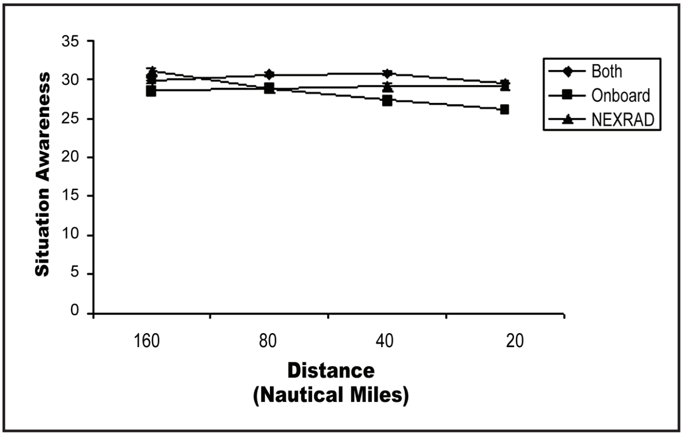


Figure 6. Situation awareness as a function of pilots' rank.

Trust

We examined the effects of time pressure, role assignment, pilots' rank, weather display, and weather system on trust using a 4 x 2 x 2 x 3 x 2 mixed ANOVA. Results showed a statistically significant three-way interaction effect between time pressure, weather display, and weather system, $F(6, 132) = 9.82$, $p < .001$, partial $\eta^2 = .31$, power = 1.00. Simple effect follow-ups showed that when the NEXRAD system was not displaying upcoming weather information at the specified waypoint, pilots' trust in the NEXRAD system significantly decreased as pilots grew closer to the potential weather event, $F(3, 141) = 20.49$, $p < .001$, partial $\eta^2 = .30$, power = 1.00 (see Figure 7).

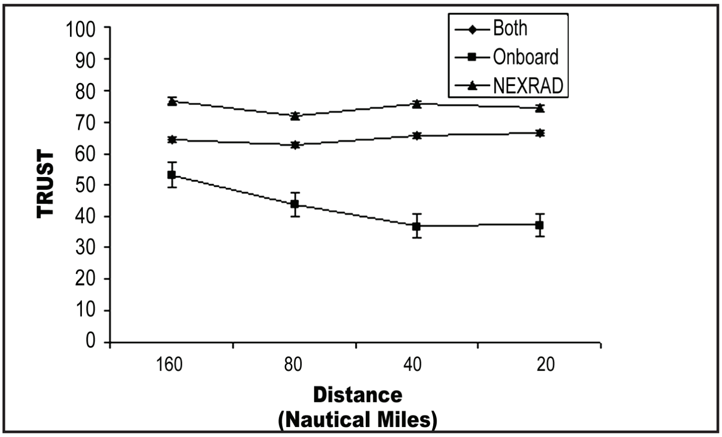


Figure 7. Pilots' trust on the NEXRAD system as a function of distance from the weather event and weather display.

Similarly, when the onboard system was not displaying upcoming weather information at the specified waypoint pilots' trust in the onboard system significantly decreased as pilots grew closer to the potential weather event, $F(3, 141) = 5.37$, $p < .01$, partial $\eta^2 = .10$, power = .93. However, when both systems agreed, pilots' trust in the onboard system significantly increased as pilots grew closer to the potential weather event, $F(3, 141) = 6.12$, $p < .01$, partial $\eta^2 = .12$, power = .96. Also when only the onboard system displayed information at the specified waypoint, pilots' trust in the onboard system significantly increased as pilots grew closer to the potential weather event, $F(3, 141) = 3.49$, $p < .05$, partial $\eta^2 = .07$, power = .77 (see Figure 8).

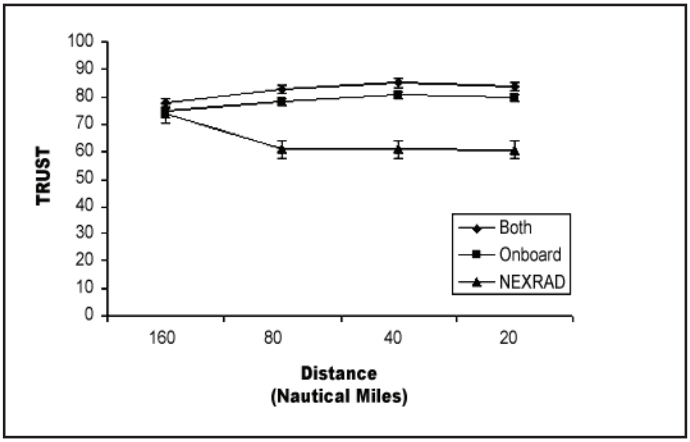


Figure 8. Pilots' trust on the onboard system as a function of distance from the weather event and weather display.

Discussion

Workload

The results partially supported our hypothesis about workload. Pilots' workload significantly increased as they approached the weather event. This finding was consistent with previous literature suggesting that distance from an upcoming weather event plays an important role on pilots' deviation decisions (Sly & Hartmann, 1999). A potential contributing factor to this workload increase could be the added time stress associated with making an accurate deviation decision as the aircraft approached the weather threat. Research suggests that time pressure is a significant contributor and indicator of workload (Hart & Staveland, 1988). Another contributing factor to this workload increase could be pilots' decreased situation awareness as the aircraft approached the weather threat (Svenson & Maule, 1993). Research suggests that although workload and SA are unique constructs (Endsley, 1995), they are very closely related (Selcon, Taylor, & Koritsas, 1991).

Situation Awareness

Results partially supported our SA hypothesis. Consistent with previous research (Jentsch et al., 1999), Captains reported lower SA than FOs. Although team members worked together to make their deviation decisions, the Captain was ultimately responsible for every decision made. This added responsibility might have reduced Captains' ability to focus their full attention on the weather event. As a result, they tended to experience lower levels of SA.

When the NEXRAD system failed to provide an indication of the weather event at the specified waypoint, pilots' SA decreased as they approached the weather threat. This finding was consistent with our hypothesis, and with previous research that suggests graphically integrated displays may enhance pilots' SA (O'Brien & Wickens, 1997; Wickens, 2000). The NEXRAD system provided pilots with a broader view of the weather event that allowed them to have higher SA. When this system failed to provide such information at the specified waypoint, pilots were force to rely solely on the onboard system. The onboard system provided them with detailed information as they approached the weather threat. However, this system did not provide them with the "big picture" needed to maintain high levels of SA.

Trust

The results partially supported our trust hypothesis. Pilots trusted the onboard system more than the NEXRAD system, particularly when these systems displayed conflicting information as pilots' approached the weather threat. This finding was consistent with previous research that suggests conflicting information affects pilots' ability to use it effectively (Pritchett, 1998). This finding was also consistent with research that suggests pilots trust reliable systems significantly more than unreliable systems (Gupta et al., 2002).

The effect of distance on trust did not support our hypothesis. We expected participants' trust in both systems to increase as they approached the weather event. Although trust in the onboard system increased as pilots grew closer to the weather event, the increase occurred only when the onboard system displayed the potential weather event at the specified waypoint. Additionally, pilot trust in the NEXRAD system never increased as pilots came closer to the potential weather event. In fact, the results suggest that in some instances, pilots' trust in the weather systems might have actually decreased as they approached a potential weather threat. When the weather systems presented conflicting information, pilots' trust in the weather system that did not display weather information at the specified waypoint significantly decreased as the pilots approached the event. When interacting with weather, pilots expect weather display reliability and agreement between displays to increase as the airplane approaches the event. If display agreement does not increase as a function of distance, pilots are forced to comply with one display and ignore the conflicting information presented on the other display. Because pilots have a responsibility to passenger safety, they typically choose to comply with the system that presents weather information. As a result, their trust in the system that does not present weather information may degrade as pilots travel closer to the event.

The findings for trust are important because system trust affects both reliance on and compliance with automated systems. As tasks become more complex, such as during weather threats, pilots need to adequately trust automated systems to be able to take full advantage of their capabilities. However, prior research suggests that humans do not always adequately map their levels of reliance and compliance to the capabilities of automated systems (Parasuraman & Riley, 1997). One of the main purposes of weather systems is to monitor the environment constantly to detect dangerous conditions. Trust, in this case, can serve to facilitate reliance, thereby reducing the workload associated with monitoring performance. Another important function of weather systems is to attract pilots' attention and sometimes suggest a course of action when problems arise. Trust, in this case, may mediate the extent to which pilots comply with the system.

Limitations and Future Research

Some elements of our experimental paradigm deserve further investigation. The major limitation of our study was the tradeoff between ecological validity and experimental control, which is common in most laboratory settings. We did not incorporate all the variables that might affect pilots' workload, SA, and trust during weather threats, such as the information provided by air traffic controllers, flight service specialists, and the airline's dispatcher. In addition, our study lacked the dynamism of real-life flight operations. Due to programming restrictions, we were unable to simulate a dynamic onboard system and present weather information throughout the entire experiment. Consequently, we were unable to assess the advantages that the onboard system would provide over the NEXRAD system as pilots flew through the different weather threats. However, the advantages attributed to the dynamic nature of the onboard system in comparison to the static nature of the NEXRAD system were not the focus of this study.

Another limitation of our study was that we did not permit pilots to deviate from the flight path. The reason why we did this was that we would not otherwise have been able to assess the effects of time pressure. For instance, if pilots decided to deviate from the flight path and avoid the weather threat 160 nm away from it, we would not have been able to assess their workload, SA, and trust at 80 nm 40 nm and 20 nm away from the weather threat.

One potential limitation of this study was the small sample size. Due to budget constraints, we were only able to collect data from 24 commercial airline pilots. Nevertheless, we did not consider this a detrimental issue for two main reasons. First, we recruited pilots from a variety of different airlines to be able to maximize the generalizability of the results. Second, the fact that we found statistically significant effects that partially supported our hypotheses with adequate effect size and power showed evidence of sufficient sample size.

One last limitation of this study is that we instructed pilots to pause the simulated flights after answering the deviation decision questions so that they could complete the NASA TLX, SART, and trust questionnaire. This could have been a limitation of the study because pilots might have had a difficult time getting back into character. However, we did not consider this a major issue because we collected data from highly experienced and trained professionals. In addition to this, most of the pilots who participated in this study had previously participated in similar studies. Future research should address the limitations of this study using a more ecologically valid design.

Conclusions

In spite of the limitations of this study, our findings have important implications for the field of commercial aviation. Consistent with previous work (Jentsch et al., 1999), our findings suggested that airlines should consider a change in role assignment philosophy. Due to the added workload associated with being ultimately responsible for the flight, Captains may benefit from allowing FOs to operate the primary flight controls. In this manner, Captains can concentrate on sources of task-critical information required to make strategic decisions. This may be particularly important during emergency situations, such as those created by weather hazards.

Airlines should also consider encouraging pilots to make deviation decisions around weather events as soon as they notice them. Although making maneuvers earlier may conflict with airline guidelines concerning fuel use and passenger comfort, it will likely increase overall safety. As pilots approach potential weather threats, workload may increase and SA may degrade. High pilot workload and low SA may hinder pilots' ability to make accurate deviation decisions.

Our findings showed support for the added benefit of providing pilots with broader information regarding the potential weather threat using the NEXRAD display. As pilots approached the weather events, their level of SA decreased when the NEXRAD system did not display weather information at the specified waypoint. This finding raises an important issue. Although the onboard system is more reliable because it presents weather information in real time, this system is limited in scope. Therefore, it is necessary to complement the onboard information with a broader view of the potential weather event.

The problem with this approach is that because the NEXRAD information is not presented in real time, it may conflict with the onboard system. Consequently, pilots may distrust the NEXRAD information, and still over rely on the onboard system. Future research efforts should explore improvements to data link technology so that NEXRAD information can be presented in real time. Such technological improvements may increase the reliability, believability, and dependability of the NEXRAD system, and ultimately avoid the distrust associated with conflicting information.

Acknowledgement

The work described in this article was supported by a research grant from NASA Langley Research Center (NAG-1-03061). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

References

Bliss, J. P., Fallon, C. K., Bustamante, E. A., & Bailey, W. R. (2005). Weather deviation decisions by air transport crews during simulated flight. In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society (pp. 140-144).* Santa Monica, CA: Human Factors and Ergonomics Society.

Bliss, J. P., Jeans, S. M., & Prioux, H. J. (1996). Dual-task performance as a function of individual alarm validity and alarm system reliability information. In *Proceedings of the IEA/HFES Congress* (pp. 1237-1241). Santa Monica, CA: International Ergonomics Association/Human Factors and Ergonomics Society.

Borg, G. (1978). Subjective aspects of physical and mental workload. *Ergonomics, 21*(3), 215-220.

Boyer, B., Campbell, M., May, P., Merwin, D., & Wickens, C. D. (1995). Three dimensional displays for terrain and weather awareness in the National Airspace System. In *Proceedings of the 39th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 6-10). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R. (1999). Situation awareness in aviation systems. In D. J. Garland, J. A. Wise, & V. D. Hopkin (Eds.), *Handbook of Aviation Human Factors* (pp. 257-276). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2000). Direct measurement of situation awareness: validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: LEA.

Endsley, M., & Rogers, M. D. (1996). Attention distribution and situation awareness in air traffic control. In *Proceedings of The Human Factors and Ergonomic Society 40th Annual Meeting* (pp. 82-85). Santa Monica, CA: Human Factors and Ergonomics Society.

Gupta, N., Bisantz, A. M., & Singh, T. (2002). The effects of adverse condition warning system characteristics on driver performance: An investigation of alarm signal type and threshold level. *Behavior and Information Technology, 21*(4), 235-248.

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Human Factors, 26*(5), 545-556.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier.

Isaminger, M., & Proseus, E. (2000). Analysis of the integrated terminal weather system (ITWS) 5-NM product suite. In *Proceedings of the 9th Conference on Aviation, Range, and Aerospace Meteorology*. Orlando, FL: American Meteorological Society.

Jentsch, F., Barnett, J., Bowers, C. A., & Salas, E. (1999). Who is flying this plane anyway? What mishaps tell us about crew member role assignment and air crew situation awareness. *Human Factors, 41*(1), 1-14.

Jones, D. G. (2000). Subjective measures of situation awareness. In M.R. Endsley & D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 113-126). London: Lawrence Erlbaum Associates.

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space, and Environmental Medicine, 67*(6), 507-512.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46,* 50-80.

Lindholm, J.M. (1999). Weather information presentation. In D.J. Garland, J.A. Wise, & V.D. Hopkin (Eds.), *Handbook of aviation human factors*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., Jr., & Wherry, R.J. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies*. (Technical Report 851). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Meshkati, N., Hancock, P., & Rahimi, M. (1992). Techniques in mental workload assessment. In J. Wilson & E. Corlett (Eds.), *Evaluation of Human Work. A Practical Ergonomics Methodology* (pp. 605–627). London: Taylor & Francis.

O'Brien, J.V. & Wickens, C.D. (1997). Free flight cockpit displays of traffic and weather: Effects of dimensionality and data base integration. In *Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society* (pp.18-27). Santa Monica, CA: Human Factors and Ergonomics Society.

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

Pritchett, A. R. (1998). Pilot situation awareness and alerting system commands. In *Proceedings of the Advances in Aviation Safety Conference & Exposition* (pp. 28-38). Daytona Beach, FL: Society of Automotive Engineers.

Sarter, N., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37*, 519.

Selcon, S. J., Taylor, R. M., & Koritsas, E. (1991). Workload or situational awareness?: TLX ve. SART for aerospace systems design and evalution. In *Proceedings of the Human Factors and Ergonomics Society 35th Annual Meeting* (pp. 62-66). Santa Monica, CA: Human Factors and Ergonomics Society.

Sherman, P.J. (2003). Appling crew resource management theory and methods to the operational environment. In P . S. Tsang & M . A. Vidulich, *Principles and Practice of Aviation Psychology* (pp. 473-506). Mahwah, NJ: Lawrence Erlbaum Associates.

Sly, J., & Hartmann, G. (1999). *Weather avoidance using route optimization as a decision aid: An AWIN topical study*. Phase I Final Report. Honeywell Technology Center. Minneapolis, MN.

Stokes, A., & Wickens, C.D. (1988). Aviation displays. In E.L. Wiener & D.C. Nagel (Eds.), *Human Factors in Aviation* (pp. 157-187). San Diego, CA: Academic Press.

Svenson, O., & Maule, A. J. (1993). *Human judgment and decision making*. New York: Plenum.
Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew system design. In *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478*. Seuilly-sur Seine: NATO AGARD.
Tsang, P., & Wilson, G. F. (1997). Mental workload. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics (pp 417-488)*. New York, NY: John Wiley & Sons.
Wickens, C.D. (2000). Human factors in vector map design: The importance of task display dependence. *Journal of Navigation, 53*, 1-14.
Wickens, C.D., Kroft, P., & Yeh, M. (2000). Database overlay in electronic map design. In *Proceedings of the IEA 2000/HFES 2000 Congress*. Santa Monica, CA: Human Factors and Ergonomics Society.
Wiener, E.L., & Nagel, D.C. (1988). *Human factors in aviation*. New York: Academic Press.
Wierwille, W. W. & Eggemeier, F. T. (1993). Recommendations for mental workload measurements in a test and evaluation environment. *Human Factors, 35*, 263 - 281.
Yeh, Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30*(1), 111-120.

International Journal of Applied Aviation Studies, 5, Number 2

Developmental Papers

Decision Making and Judgment

John E. Bertrand
Department of Aerospace, Box 67
Middle Tennessee State University
Murfreesboro, TN 37130

Abstract

Good decision making and the exercise of good judgment have been issues of interest to the aviation profession and the Federal Aviation Administration (FAA) for many years. The traditional step-by-step model of decision making has had a long life, but has failed to serve pilots in the real world very well. Aviation decisions are risky, time constrained, fraught with variables, and seldom present the decision maker, the pilot, with anything approaching complete information. As such, they are very different from the leisurely decisions usually associated with the step-by-step or "classical" method. This article examines the research and experience of the last 20 years that examines real world decision making by aviators and others in similarly placed positions. It examines two models of decision making which are much better suited to aviation and comments on their potential for wider application in the profession.

Many authorities believe that judgment, based on integrated decision-making, driven by excellent situational awareness and knowledge of what lies ahead, is the single most important group of skills for safe and efficient pilot performance. Some accidents are failures of mechanical components; some are based on a gap in the pilot's knowledge base that led to an accident of ignorance. But most are adjudged pilot error, another way of saying that the pilot had a failure of judgment. Decision making is the intake, processing, and management of data brought in by the senses. It is, in effect, a sub-set of learning, the most critical sub-set in avia-

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

tion. It is the purpose of this paper to examine the historically accepted means of modelling decision making, usually called Aeronautical Decision Making (ADM), and contrast it to models that have emerged in the last twenty years.

Judgment is a slippery concept. Only its outcomes can be seen, and it can only be taught successfully via indirect means. In the *Aviation Instructor's Handbook* (2003), the FAA addresses the issue in an eleven-page essay. In the *Practical Test Standards* (1991), the issue is not addressed, except tangentially in passing. Therefore, while the FAA wants instructors to be aware of the concept of aeronautical decision making (ADM) and some of the issues that go with it, there is no direct means of evaluating it on the strength of the stimulus-response (Skinner, 1938, 1969) world of present instruction. Part 141 schools (which are more often aimed at those who want to be professional aviators) probably address this set of issues more completely than Part 61 schools (which are more often aimed at those who wish to be private pilots) because the texts used in Part 141 formal classes (such as the Jeppersen [2002] textbook) give some attention to ADM.

The Traditional Decision Making Model

The "classical" decision-making model is exactly the one described in the *Aviation Instructor's Handbook* (FAA, 2003). It has been stated over the years with anywhere from three to seven steps, but it is usually first credited to Drucker (1954), the famed business management scholar and author.

The basic steps of the model are:

1. Identify the problem
2. Develop alternatives
3. Compare potential advantages and disadvantages of each
4. Choose an alternative
5. Evaluate the outcome

This set of steps was originally posited as a business decision model and has been adapted over the years with a steadily growing number of accompanying checklists to assist pilots in making decisions. The best known of these is that of Berlin, Gruber, Jensen et al. (1982) that listed "The Five Hazardous Attitudes." Following the five hazardous attitudes are the IMSAFE checklist, the Pitfalls checklist, the DECIDE model of decision making, the PAVE decision making model, the SDRV decision making model, and the concepts of risk, stress, and workload management. All but two of these are still in the *Aviation Instructor's Handbook* (FAA, 2003).

Though the *Aviation Instructor's Handbook* (FAA, 2003) does not recommend conflicting decision-making models side-by-side, a student would have to take account of quite a body of criteria before making any important decision in an airplane. The IMSAFE, hazardous attitude, risk management, and pitfalls checklists are apparently criteria pilots should check in with frequently while flying to make sure that they or their operations have not become dangerous. They should also engage in risk, workload, and stress management while flying. This is in itself a workload, of course (Deitch, 2001).

Therein lies the problem that has attracted the attention of scholars in many fields who have become dissatisfied with the traditional model. All these are normative techniques; that is, they are prescriptive. Deitch (2001) called them “cognitive prompts,” since they serve primarily to merely prompt trains of thought. A primary factor of such models is that they are designed to either cancel the flight altogether or to cause pilots to decide, in advance, what will be done in a given set of circumstances. Risk management, as an example, urges pilots to decide in advance what measures will be taken in the event of an engine failure, a fire, becoming lost, shooting an unsuccessful approach, and so on. The rub lies in that events overtake even the most cumulative, on-going, Drucker-style (1954) process.

Such normative models have their place, of course, but they are impractical to the needs of many occupations and professions wherein decisions are made under pressure. These include military battlefield management, fire fighting, search and rescue, and a number of others in addition to aviation. All the ways the original Drucker (1954) model have been adapted to various venues carry the original set of assumptions with them. The bare list of steps may fairly be expanded to illustrate what one might really have to do to make the model work as follows [underlines added by this author] (Kaemph & Militelo, 1992).

1. Specify *all* the relevant features of the task.
2. Identify the *full* range of options.
3. Identify the *key* evaluation dimensions.
4. Identify *weights* for each dimension.
5. Rate *each* option on each dimension.
6. *Tabulate* the results.
7. Select the *best* option.
8. Evaluate the outcomes.

Obviously, this set of steps might be a little difficult when one is flying in complete darkness alternating with lightning flashes, trying to maintain the aircraft on the glide slope and localizer, deal with the tower’s second request to know if one is inside the outer marker, deal with the objects flying around the cockpit due to turbulence, and ignore the cries of one’s passengers. In such a situation, the decision to elect a missed approach is far from a sequential process. It is based on fear, the amount of fuel remaining, and a number of other factors that are *not* reviewed one at a time. Moreover, as the literature points out, the decision reached will probably not be very rational; it will be intuitive and difficult to reconstruct in logical terms.

The assumptions buried in the classical process have been laid out by Hutchins (1996). They are that:

1. decision makers have sufficient time to generate options, conduct option assessment, and select a course of action
2. the consequences of an incorrect response are not immediately severe
3. decisions are reached with input from others, and
4. the workload is manageable.

Unfortunately, flying, as Hutchins points out, is characterized by high risk, uncertainty, information ambiguity, high workload, and task complexity.

In the last fifteen years or so, the issue of decision making in real-life milieus has been understood differently. This is the result of observational, survey, and interview studies in various fields wherein decisions are made under pressure, while the classical model still assumes the luxury of addressing the context in which the decision is made, using all information, facts, and data. The implicit assumption of the classical model is that actually making the decision will be automatic and unremarkable once enough understanding has been brought to the problem; a view that is considered no longer sufficient to the needs of aviation since flying often does not generate enough time or information.

More recently, interest has shifted to recognize that dynamic situations seldom provide complete information and to instead address the decision maker. Adherents to the Drucker (1954) model might object that the classical decision making process allows for incomplete information. However, the sorts of activities posited here anticipate a complex, shifting, data-poor scenario with diminishing resources that would send the traditional adherents of the process, MBAs, and other business professionals, quickly to the point of paralysis. Decision makers who are fire fighters, pilots, warriors, and so on do not have the luxury of pondering too far. They will decide how to react, for better or worse, and they always know that to make no decision is a decision in itself.

Two unfortunate incidents in 1992 and 1994, the USS Stark and Vincennes disasters, accelerated the trend. Today, researchers are interested in examining the decisions of naval surface ship commanders, tank platoon leaders, fire commanders, offshore oil installation managers, and infantry officers, as well as pilots and others (Lipshitz, Klein, Orasanu, & Salas, 2000). All these are people who make life-or-death decisions based on a lack of complete information, using mental processes that decidedly do not reflect the logical and linear premises of the classical model, and wherein experience instead of intellect appears to be the best predictor of success.

The Naturalistic Decision Making Model

So how then might the results of more recent thinking and research be characterized?

> The naturalistic approach to decision research takes, as its starting point, the way people actually make decisions in real-world environments, as revealed in interviews, observation, and contextually realistic experimentation. It does not start with a mathematical or logical model of how decisions ought to be made, nor does it typically compare behavior in artificial laboratory tasks to such models. However, there is more agreement about the starting point of naturalistic research than its destination. (Cohen & Freeman, n.d., p. 1)

Emphasis is therefore placed on the decision maker, how s/he gathers, synthesizes, analyzes, and responds to incomplete data, shifting situations, and urgent consequences (Lipshitz et al., 2000). The reasons for this in aviation are:

1) the relative complexity of the decision milieu,
2) the on-going nature of the decision process, and
3) the interaction of the next decision with all those that remain to be made.

All decisions, made classically or naturalistically, are conditioned by circumstances. By definition, naturalistic decision making (NDM), is the response of choice when the following conditions (as they certainly do in flying) apply:

1) uncertain dynamic environments
2) multiple-event feedback loops
3) meaningful consequences
4) multiple goals
5) time constraints
6) complexity of decisions
7) multiple players
8) large quantity of information
9) level of expertise required

To respond to these conditions, nearly all NDM models have a few points in common. First, all NDM models are open to “satisficing”, a combination of the words “satisfactory” and “suffice” which embody the notion that the decision should be driven by the first alternative that is “good enough”. There is little drive in most NDM models to find the “best” or “perfect” alternative, considering that the situation includes high risk, time pressures, and uncertain conditions. Second, the decision making process does not agonize over non-solutions (compatibility testing). Third, biases, personal experience, and expertise are large drivers towards the decision selected. Fourth, no clear-cut algorithmic, classical approach is possible. The problem by definition contains an enormous number of interrelated variables and situations that the decision maker must decode and interpret. Lastly, this is done by a process of mental imagery, one that may be highly idiosyncratic.

NDM has been of interest to researchers and others since the late 1980s. The NDM general framework discussed above is credited to a conference in Dayton, Ohio sponsored by the Army Research Institute in 1989. In general, the outcomes of this conference recognized not only time and risk constraints, but also the importance of examining people who exhibited expertise (and how they obtained it), and the acceptance of the notion that real world decision making often made situation assessment more important than decision methodology. That is, the way people sized up situations seemed more critical that the way they selected courses of action (Klein, 1993). A series of conferences every three years or so has further elaborated the research framework.

In general, what have emerged are themes on two generally recognized models. The first of these is generally referred to as Recognition-Primed Decision Making (RPD) and is credited to Klein (1993). In its current form, the RPD model has three variations. In the simplest form, the decision maker sizes up a situation and responds with whatever seems the best option based on expertise. The driver is experience, which allows decision makers to formulate typical case scenarios, wherein certain responses are usually adequate and require little further examination. The second form applies when the situation is unclear. Here, the decision maker will rely on story building to mentally simulate events leading up to the situation as observed (Klein & Crandall, 1995; Pennington & Hastie, 1993). In the third variation on this theme, the decision maker mentally simulates several proposed actions and examines them for unintended consequences that might be unacceptable.

In variation one, a pilot might become aware of an isolated thunderstorm in the path of travel and make a standard style judgment based on winds aloft, course, fuel remaining, traffic, passenger comfort, urgency of schedule and so on. After a holistic consideration, the pilot then decides to alter a number of degrees off course to pass a number of miles away from the danger. Here, unless a radical new variable intrudes, the decision, while hardly reflective of a classical, linear process, still reflects the first option identified.

In variation two, the pilot may find that the forecast is incorrect. The forecast line of thunderstorms is moving faster than anticipated with concurrent turbulence and winds aloft changes in advance of the frontal boundary. Now the situation is unclear. The typical GA pilot in this position has anything but full information, is faced with a changing situation, and has no choice but to adopt a new plan. Here, the pilot calls on knowledge of frontal and thunderstorm behavior, a mental overview of location and speed of the weather, all the variables listed in scenario one, the capabilities of the airplane, and a risk assumption comfort level to arrive at a plan that offers the best probability of success. A combination of knowledge and experience is used to match the risk comfort level with probable outcomes of the plan.

In variation three, a pilot might find (as is common) that unforecast icing conditions are being entered. Now, a problem exists that must be addressed much more immediately, cannot be solved with as trivial a solution as a 20-degree course change (as in the thunderstorm scenario), and for which information is both more scarce and more diffuse. (Weather information is typically good for thunderstorms but pretty poor for icing.) In this case, the decision maker must create and critique several scenarios that lead to a good enough solution. Probably nothing is going to make the ice go away. The “satisfying” solution is one where the trip is completed without unacceptable danger. In many cases, the only acceptable solution may be mission failure in the form of an immediate precautionary landing. This third variation relies on the construction of a narrative by the pilot that explicates and delineates the problem and suggests alternative solutions.

The three variations depend heavily on expertise. The RPD model is a good description of how experts make decisions because it differentiates between routine decisions and situations where the situation is quite unclear. In addition, it describes how expert decision makers work forward, from existing conditions, rather than backwards, from goal states, when the stakes are high. Novices, on the other hand, usually rely on backward chained reasoning or stated, context free rules not directly linked to the situation. An example is the directive to always climb when entering icing conditions. That will work in the majority of cases, unless these icing conditions are an exception (for example, entry into super-cooled droplets), if the airplane has enough surplus power to climb successfully, and the warmer air is not above the service ceiling of the airplane, and so on.

This model of decision making appears to apply only in certain conditions, as previously stated. Lipshitz et al. (2000) enumerated them as follows: 1) when the decision maker has reasonable experience to draw on, 2) when there is time and risk pressure, and 3) when there is uncertainty and/or ill-defined goals, which provide the decision maker a number of degrees of freedom in terms of outcome(s). The RPD model is less likely to be used with highly combinational problems, where justifications are necessary, and in cases where the views of different stakeholders must be taken into account. Given conditions wherein the RPD model is the decision making paradigm of choice, decision makers who are proficient in the practice of RPD usually perform acceptably and often exceptionally through the effective use of pattern matching, forward-directed reasoning, and narrative construction.

However, uncertainty can become so high that the pattern matching strengths of RPD fail (Cohen, Freeman, & Wolf, 1996). Lack of experience or inability to combine available information into a proposed solution may lead to paralysis or a reversion to “rule bound” reasoning. Cohen and his colleagues perceived that RPD relies primarily on pattern matching. Their contribution addresses what happens when recognition fails. If stakes are high and time permits, decision makers revert to assumption-based reasoning, which, as elaborated in the model, consists of meta-cognitive processes of critical thinking by which decision makers identify and correct gaps in situational awareness and action plans owing to incomplete or conflicting information, inconsistent goals, and unwarranted assumptions. That is, decision makers not only think about the decision; they also ask themselves if they are thinking about it correctly or if some other point of view or lens could reveal the problem more accurately. They called this Recognition/Metacognition (R/M).

For situations of high uncertainty where stakes are high, Cohen and his associates (Cohen & Freeman, 1997) developed a generic procedure based on R/M which they called STEP (story, test, evaluate and plan), which can be applied to improve performance on a decision task which requires perceptual input. STEP has been applied to quite a number of situations wherein information is scarce and risk is high, including engage/don’t engage decisions in ambiguous tactical situations, fireground command decisions concerning the allocation of fire fighters and apparatus, diffuse aviation situations, and others. While engaging in a four-step procedure is somewhat prescriptive, the intent is to allow decision makers to emulate more expert performance. The process is based on field research into the actions of real-world decision makers. It avoids the fallacy of starting out with predetermined outcomes, and it admits the possibility of error and its correction (Cohen & Freeman, n.d.).

1. Story – Even pattern matching which yields only vague recognition generates some conclusions concerning the situation. The decision maker here constructs a Story that recounts past, present, and future events consistent with as much pattern as can be recognized. Thus, a pilot may ask… "How did I get in this mess? What are the elements of the situation that control what is going on and what is my present condition? If things develop as I think they will, what is going to happen?" Diffuse and amorphous information is thus arranged into a comprehensible narrative that examines the future as well as present.

2. Test – Stories are used to test the plausibility of initial assessments by comparison of implications and expectations derived from them, with what is known or observed about the situation. When evidence appears to conflict with an assessment, stories are revised to incorporate all available information into the most complete and plausible account possible. The intent is to identify and fix gaps in stories owing to incomplete data or unwarranted underlying assumptions.

3. Evaluate – In this phase, the decision maker uses a devil's advocate technique in which an outside force repeatedly insists that the current story is wrong and asks for an explanation. When amended stories require too many unconfirmable assumptions, the decision maker is directed to return to step one.

4. Plan – Here, the decision maker constructs a back-up best plan that is available on short notice, qualified by its known strengths and weaknesses. This is used as a hedge against the possibility that the current best plan is wrong. In doing this, the process admits the possibility of error and its correction.

At the conclusion of this process, ambiguity should be lessened and confidence in the plan of action strengthened. That is, augmented Situational Awareness leads to greater confidence in the Course of Action (Orasanu & Martin, 1998).

A simple example of the process might sound something like the following:

Story: Hmmm… I smell hot metal, I think. Let's see; everything looks normal. So did I spill some oil when I checked it? No, I don't think so. Is this plane a leaker? No, not that either. But I bet I am smelling oil baking on hot metal and the only hot thing around here is the engine. I know I checked the oil before I left and there was enough, so I should not be smelling baking engine, but it is not going away. If I have an oil quantity problem, this engine is going to shut down in a matter of time, hard to tell how much time. And it may shut down catastrophically and put any remaining oil on the windshield as it dies. Hmmm… I am 85 miles from my destination at 9000 feet. This plane will glide about 13 miles from this altitude. The terrain below is broken farm field and forest. So… what to do? I think I will keep a real sharp eye on the oil pressure and temperature gauges and look for any variation at all. Any change will probably indicate that the engine is in final failure and prone to lock up. If that happens, I will take a precautionary landing as soon as possible. The alternate airport I filed for has better weather than my destination and is a little closer, so I think I will land there anyway just in case. There… that's a plan.

Test: So… is there anything wrong with this plan? Well… if we are considering a precautionary landing, let's just hit the "nearest" button on the GPS to make sure there are some airports around here. OK… there are, good. And is a climb a good idea here? Might give me a longer glide. On the other hand, I will slow down as I climb and end up running the engine longer. No, I think 13 miles glide is enough for this terrain. I can surely find a cornfield or a local airport in that distance if things really go bad. Is there anything else I haven't thought of? No? So OK.

Evaluate: But what if I'm wrong? What if I've got the whole situation wrong? What if it is an electrical smell? Do I need to go back and plan this for a fire in the cockpit? How do I know that is an oil smell? Sniff, sniff. Well, I swear it is. I am going to go with the oil leak plan. I'll go on to the planned landing.

Alternate Plan: Oh, man. What if it is electrical? Well, let's see about plan B. If I see any smoke, I will cut off the master switch and try to isolate the problem by cutting one thing at a time back on. I'll check the breakers. And I will divert immediately to the closest airport. Is the extinguisher where it ought to be? OK. That's plan B, just in case I'm wrong.

Can this sort of reasoning be learned in other than life-or-death situations? Cohen, Freeman, and Thompson (1998) determined in their study that it could. In their study, the control group received training of a more conventional kind. They found that any training at all increased the subjects' confidence in their abilities, but that only the meta-recognition group showed a statistically significant concurrent increase in performance.

Craig (2000) recounted what he believes is the best example of NDM training available to civilians, the full-motion simulators employed by the airlines and such companies as Flight Safety and Simcom. Under Special Federal Aviation Regulation (SFAR) 58, air carriers and approved others can train pilots using pilot proficiency and decision making capability rather than maneuvers for pilot evaluation. This training is legally logged by pilots who are already required to be trained by Part 121 (regular air carrier) or Part 135 (on-demand charter), and the training must take place on approved simulators. This is called the Advanced Qualification Program (AQP).

Each approved curriculum has three parts: indoctrination, qualification, and continuing qualification, depending on the credentials and experience of the participant. No number of hours is of consequence. Very experienced pilots sometimes fail and must take additional training. Some talented low time pilots excel. Performance under pressure is what counts. Pilots are compelled to manage information, utilize resources, work around problems, make decisions, and ultimately, arrive or crash. Curricula are scenario based on Line Oriented Flight Training (LOFT), and performance is tested via Line Oriented Evaluation (LOE), which consists of real time, simulated flights from one airport to another. Unusual occurrences are guaranteed.

Craig's (1998) dissertation examined what would happen if GA, lightplane pilots were exposed to LOFT training. He found that qualitative pilot performance was much improved on the arrive-or-crash posttest scenario, which he posed to more than a hundred pilots on a Frasca 141 simulator. All his subjects were licensed IFR rated pilots, who averaged about 350 flight hours. In his pretest (on the Frasca 141), after presented with a scenario wherein subjects were forced to contend with a slowly failing airplane and had to go to an alternate airport, 68% crashed or had a forced landing in IFR conditions. Following LOFT training, this number declined by about half to 34%. Given the level of intervention that his treatment represented, this was a startling improvement.

Summary

For nearly twenty years, those interested in decision making have realized that many situations do not lend themselves to lengthy reflective or linear processes. Emphasis in research has shifted therefore away from analyzing the situation or context and toward examining the decision maker, particularly the processes by which experts make decisions. A number of conclusions may be drawn from this work.

1. The performance gap between experts and novices who have completed traditional training is very wide. Experts perform significantly better in time constrained, risky, diffuse situations than even novices who have received extensive, lengthy traditional training.
2. The driver of this large difference in performance appears to be experience and the development of decision making expertise.
3. The more diffuse and ambiguous the situation, the greater the gap between experts and novices.
4. Traditional, linear situational analysis cannot account for the variables in diffuse situations and tends to break down.
5. Experts use pattern matching skills to holistically analyze ambiguous situations and arrive at decisions that offer a greater probability of success.
6. Experts far exceed what might be predicted for their performance based on available knowledge. Pattern matching skills that are developed over time by application allow for adequate results, error identification and correction, and conclusions that reflect complex mental operations on both a conscious and unconscious level. All these processes taken together are called "expertise."
7. These skills can be learned to a significantly greater level in safe milieus, using simulators, certain types of classroom training (Craig, 1998), and non-traditional, scenario based, dual instruction.

Recommendations

As interest continues to build in scenario based instruction both for ab initio training and transition training, it is clear that the development of judgment and expertise should be recognized and addressed directly with a consistent model. Those who receive this sort of training perform much better than those who receive traditional training and very often, in excess of what their experience predicts.

The results of the SAFER program, a year and a half long NASA- and FAA-sponsored ab initio training program at Middle Tennessee State University, are not yet published but point the way. In this study, absolute beginners were offered combined private/instrument training, in all-glass cockpits, using strongly scenario based training methods offered by specially trained instructors. Particular attention was paid in each lesson to the building of judgment and expertise. The results are startling. First, most of these pilots received the private license with instrument rating after about 60 flight and 25 simulator (in a full vision sim) hours. More importantly, their decision making skills and ability to cope with constantly shifting events approximate pilots with 350 to 400 hours.

It seems clear that the development of expertise can be accelerated and that the use of good judgment can be assisted via models which make its creation more explicit. Training which more closely approximates LOFT and LOE, as the SAFER study does, holds a great deal of promise. The FAA is working with Middle Tennessee State University, University of North Dakota, and Embry-Riddle Aeronautical University to address these issues as part of the FAA Industry Training Standards (FITS) program, and the work shows promise. However, those interested in naturalistic decision making should become involved in the creation of the new curricula, and the proven models of this area of interest should be included.

References

Berlin, J., Gruber, C., Holmes, P., Jensen, I., Lau, R., Mills, J. & O'Kane, J. (1982). *Pilot judgment training and evaluation (DOT/FAA/CT 82/56).* Daytona Beach, FL: Embry Riddle Aeronautical University.

Cohen, M. & Freeman, J. (n.d.). *Improving critical thinking.* Retrieved April 2005, from the Cognitive Technologies Inc. Web site: http://www.cog-tech.com/papers/chapters/aberdeen/aberdeen/pdf.

Cohen, M. & Freeman, J. (1997). Understanding and enhancing critical thinking in recognition based decision making. In Finn, R. & Martin, L (Eds.) *Decision making under stress: Emerging themes and applications.* Aldershot, UK: Ashgate.

Cohen, M., Freeman, J. & Thompson, B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. In Cannon-Bowers, J. & Salas, E. (Eds.). *Decision making under stress: Implications for training and simulation.* American Psychological Association.

Cohen, M., Freeman, J. & Wolf, S. (1996). Meta-recognition in time stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*(2), 206-219.

Craig, P. (1998). *Improving pilot decision making in situations of high stakes, high stress, and time pressures.* Doctoral dissertation: Tennessee State University.

Craig, P. (2000). *Pilot in command.* New York: McGraw-Hill.

Deitch, E. (2001). *Learning to land: A qualitative examination of pre-flight and in-flight decision making processes in expert and novice aviators.* Doctoral dissertation: Virginia Polytechnic Institute and State University.

Drucker, P. (1954). *The practice of management.* New York: Harper.

Federal Aviation Administration (1991, May). *Flight Instructor Practical Test Standards for Airplane (FAA-S-8081-6A).* Washington, D.C.: Flight Standards Service.

Federal Aviation Administration (2003, July 15). *Aviation Instructor's Handbook* (FAA-H-8083-9). Washington, D.C.: U.S. Government Printing Office.

Hutchins, S. (1996). *Principles for intelligent decision aiding:* Technical report 1718. Retrieved April 2005,:from, http://spawar.navy.mil/sti/publications /pubs/tr/1718/trl11718.pdf

Jeppersen Sanderson, Inc. (2002). *Private pilot manual.* Englewood, CO: Jeppersen Sanderson Training Products.

Kaemph, G. & Militelo, L. (1992). *Decision making in emergency first offshore management conference.* Emergency Command Responsibilities Collected Papers. Aberdeen, Scotland.

Klein, G. (1993). A recognition primed decision (RPD) model of rapid decision making. In Klein, G., Orasanu, J., Calderwood, R. & Zsambok, C. (Eds.), *Decision making in action: Models and methods.* Norwood, CT: Ablex Publishing Corp.

Klein, G. & Crandall, B. (1995). The role of mental simulation in naturalistic decision making. In Hancock, P., Caird, J., & Vincente, K. (Eds.) *Local applications of the ecological approach to human-machine systems.* Hillsdale, NJ: Erlbaum.

Lipshitz, R., Klein, G., Orasanu, J. & Salas, E. (2000). *Taking stock of naturalistic decision making.* Retrieved April 2005,:from the Center for the Study of Organizations & Human Resource Management Web Site: http://organizations.haifa.ac.il /html/html_eng/raanan%20-%20taking.doc

Orasanu, J. & Martin, L. (1998) Errors in Aviation Decision Making: A Factor in Accidents and Incidents. *In Proceedings: HESSD,* 1998, 100-107. Retrieved April 2005,:from http://www.dcs.gla.ac.uk/~johnson/papers/ seattle_hessd/judithlynne-p.pdf

Pennington, N. & Hastie, R. (1993). A theory of explanation-based decision making. In Klein, G., Orasanu, J., Calderwood, R. & Zsambok, C. (Eds.), *Decision making in action: Models and methods.* Norwood, CT: Ablex.

Skinner, B.F. (1938). *The behavior of organisms: An experimental analysis.* Morgantown, WV: The B.F. Skinner Foundation.

Skinner, B.F. (1969). *Contingencies of reinforcement.* New York: Appleton-Century-Crofts.

International Journal of Applied Aviation Studies, 5, Number 2

Book Review

Another Perspective of Aviation Education and Training

Todd P. Hubbard
Oklahoma State University

Review of
Aviation Education and Training
by Irene Henley (editor)

Irene Henley's text on aviation education and training supplements our reading list of other helpful texts on the same subject by targeting the weak areas of most collegiate aviation programs: the absence of any definable theoretical approach to learning and the lack of imaginative strategies to meet the needs of adult learners. From the first pages of the preface to the concluding summary, there is no doubt that the text beacons the attention of educators and trainers.

In an era of male-oriented aviation literature, this text is decidedly female-friendly. Out of the 11 authors, 9 contributors are female. You will also find case studies and examples drawn from experiences of female student pilots, particularly present in the chapter by Dianne Conrad and Jo Harris. Gender issues also arise in Henley's chapter on learning styles, multiple intelligences, and personality types (see Chapter 4). Outside of studies of women in collegiate aviation programs (see works by Turney and others), you'll not find a better collection of thoughts on gender-oriented aviation education and training.

Readers who have not taken courses in educational psychology, adult education, or instructional design might find the going a bit rough in Part 1 of the book. Chapters written by Bye or Henley are particularly technical, although not too complex as to mask the relevance of the material. Pilot instructors who have more experience teaching students how to fly and little experience teaching students courses in law, management, and ethics might find Part 1 unnecessarily obscure. However, my suggestion that many pilot instructors might find Part 1 difficult, might indicate the presence of a shortfall in the complete education of pilots. Have we been focusing too much on teaching pilots how to fly and not enough on preparing pilots to embrace all aspects of aviation? Perhaps Part 1 is the litmus test for educational competency among aviation educators. If you understood Part 1, you are educationally competent.

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

Those lacking in technical competency in education will still find this book useful, because Part 2 of Henley's text is a welcomed review of teaching strategies, supported by real-world examples. I found it particularly helpful that the editor used a cross-referencing technique to connect the theories and educational concepts from Part 1 with the practical use of those theories and concepts in Part 2. This tied the book together nicely.

I would recommend this book to aviation faculty members who have had little exposure to education theory. It's a start, without overwhelming a person with the hundreds of sources that fully express each theory. In fact, I would recommend that it be required reading for all aviation faculty, because of its potential to stimulate meaningful discussion among undergraduate and graduate students.

In addition, I would put this book on the required reading list for all aviation doctoral candidates. Henley's chapter on using problem-based learning is particularly fitting to doctoral candidates, because she lays out the blueprint for problem-oriented analysis, often used in quantitative and qualitative dissertations (see Chapter 11).

The text is not without some difficulties, but most of these are differences of opinion or disagreements over which theory should be supported and which theory should be ignored. At times the authors get bogged down in profound explanations of the finer points of theory, without supplying specific illustrations of the theory in the practice of aviation (see Chapters 1 & 2).

Therefore, I recommend that readers not read this text from beginning to end. Since Henley uses cross-references in Part 2, it's possible to start reading at Part 2 (see Chapter 7). The reader could use the cross-reference to Part 1 chapters as a way to review only those theories pertinent to the Part 2 teaching strategies under review.

Until this text, Australian views of aviation education and training had been trumpeted by persons such as Ross Telfer (see *The Psychology of Flight Training*, Telfer/Biggs, Iowa State Press (1988)). It is good to see that there are generations of aviation educators in Australia, particularly at the University of Western Sydney.

As a last note, I would like to comment on the physical appearance of the book. Ashgate Publishing did a good job with the presentation of the material. Immediately after the contents page and list of figures, the reader will find biographical notes introducing each author. This is followed by a very informative preface and a list of abbreviations. Thus armed, the reader is fully informed and ready to read. The marginal spacing makes reading easy and the numerous subtitles help chunk the various subordinate topics within each chapter. The dust-jacket information reinforces the preface and reaffirms the target audience as aviation practitioners. All of these feature and fixtures combine to present the reader with an intelligent treatment of aviation education and training. Nicely done.

TpH

International Journal of Applied Aviation Studies, 5, Number 2

Erratum

Automated Hover Trainer: Simulator-based Intelligent Flight Training System

John E. Stewart II and John A. Dohme

The paper "Automated Hover Trainer: Simulator-based Intelligent Flight Training System," by John E. Stewart II and John A. Dohme that appeared in the International Journal of Applied Aviation Studies, Vol. 5, No. 1, pages 25 through 40, omitted Table 5 on page 34. The Table below shows Table 5 as it should have been presented.

Table 5
Mann-Whitney U test on Unassisted Iterations Not Performed to Standard in the Aircraft as a Function of Pretraining in the Automated Hover Trainer

Test Statistic	Stationary hover	Hover taxi	Hover turns	Landing from a hover	Takeoff to a hover	Total
U	144.50	114.50	162.50	182.00	174.50	141.50
z	-2.21	-2.90	-1.80	-1.34	-1.52	-2.27
p (one-tailed) <	.02	.002	.04	.09	.07	.01

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

CONTRIBUTOR INFORMATION

Content: The International Journal of Applied Aviation Studies publishes formal papers and training development articles, which supply the reader with alternative approaches to aviation challenges. The journal was born out of the concept of mutual support. Nations finding solutions to aviation problems at home could help other nations facing the same or similar problems in their country. Some of the many areas of interest of this journal are: operational support, regulatory standards, training support, airway facilities, air traffic control, airports and logistics, aviation security, and international training.

Manuscript: Rather than follow the traditional process of supplying three copies of your manuscript, we prefer that you send your manuscript via e-mail to the editor, Beverly S. Laughead (bsl@ou.edu). We prefer that each manuscript be prepared in Microsoft Word (at present, we are using Microsoft Office 2000). All figures and tables should be prepared as tif, .bmp, or .jpg files (preferably in 300 dpi), separately from the figure or table title. Tables may also be created in Microsoft Word. All figures and tables with figure and table titles should be incorporated into the manuscript. Please provide an abstract and a running header title for your manuscript. Additionally, please include your address, telephone number, and e-mail address after the names of the authors. If you want this information withheld during publication, please advise.

The manuscript should conform to the standards of the Publication Manual of the American Psychological Association (5th ed.). Be sure to reference this document when compiling the reference portion of your manuscript. We prefer italics rather than underlined titles or journal names.

Permissions: The editor assumes that all copied material has been included by permission of the original author or has been suitably cited.

Regulations: This journal publishes original work. We assume that you have not previously had this article published, nor submitted this article to another publication. In addition, we expect that each article has met the ethical standards of the American Psychological Association or suitable alternate during the treatment of your subjects.

CONSULTING EDITORS

Consulting editors are appointed by the Editor to represent all elements of aviation. An appointment recognizes professional and academic achievement. For consideration, submit a curriculum vita or statement of interest to Beverly Laughead, Editor.