International Journal of Applied Aviation Studies

Spring 2004

REVIEW PROCESS

The Federal Aviation Administration Academy provides traceability and over-sight for each step of the *International Journal of Applied Aviation Studies (IJAAS). IJAAS* is a peer-reviewed publication, enlisting the support of an inter-national 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 Beverly Laughead by email at bsl@ou.edu, by telephone at (405) 954-1191, or by writing to the following ad-dress:

International Journal of Applied Aviation Studies
c/o Beverly Laughead
AMA-500-OU
PO Box 25082
Oklahoma City, OK 73125

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.

CONSULTING EDITORS (con't.)

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	Middle Tennessee State University, Murfreesboro, TN
James M. Hitt, II	Booz Allen Hamilton, McLean, VA
Judy Holcomb	FAA Academy, Oklahoma City, OK
Todd P. Hubbard	University of Oklahoma, Norman, OK
Michael Hynes	Hynes Aviation Services, Frederick, OK
Raymond E. King	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
Rebecca Lutte	University of Nebraska at Omaha, Omaha, NE
Beth Lyall	Research Integrations, Inc., Tempe, AZ
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	Parks College, St. Louis University, St. Louis, MO
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
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
Esa M. Rantanen	University of Illinois, Urbana-Champaign, IL
Gail Reed	Pan Am International Flight Academy, Miami, FL
Robert Ripley	Auburn University, Auburn, AL
Art Saboski	Embry-Riddle Aeronautical University, Prescott, AZ
Nadine Sarter	The Ohio State University, Columbus, OH
Robert Schlegel	University of Oklahoma, Norman, OK
Kenneth Sperry	The Boeing Company, Midwest City, OK
Alan J. Stolzer	Parks College, St. Louis University, St. Louis, MO
Craig Gerald Stroup	St Louis University, St. Louis, MO
Donald Anders Talleur	University of Illinois, Urbana-Champaign, IL
Ron Ward	FAA Academy, Oklahoma City, OK
Thomas Weitzel	Embry-Riddle Aeronautical University, Daytona Beach, FL
Michael Wiggins	Embry-Riddle Aeronautical University, Daytona Beach, FL
Peter Wolfe	Professional Aviation Board of Certification, Washington DC
John Young	Purdue University, West Lafayette, IN

PHILOSOPHY STATEMENT

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

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

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

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

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

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

B.S.L.

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

EDITOR'S NOTES
Formal Papers

Johnson, Rantanen, and Talleur describe the development of metrics of pilot performance derived from time series of various flight parameters and examine their descriptive power and sensitivity against data from two pilots with known differences in performance.

Aviation policymakers may find the Bowen and Lu article to be quite interesting. The authors present their argument that serious deficiencies exist in traditional policymaking and propose a policy research construct that they contend not only tries to address the core of policy analysis and policy evaluation, it also tries to bridge the gap between two policymaking doctrines.

Since spatial disorientation (SD) in flight is a frequently mentioned cause of accidents in military aviation, military aviators may find the Kallus and Tropper article interesting. The authors conducted a study to evaluate the efficiency of a disorientation-recovery program using the Airfox DISO flight simulator.

Those interested in improving the aviation safety culture, as well as aviation educators and trainers interested in instruction for converting fuel on board into flight time and preflight inspection instruction, will find the Dillman, Lee, and Petrin article to be of interest. Using the hypothesis that there were deficiencies in the preflight process of some students, the authors conducted a study to determine whether standardizing the process of preflighting an airplane for flight students as a group would strengthen the process of establishing the culture of safety for a given flight.

Those interested in the human factors approach to aircraft maintenance may find the Fogarty article to be of interest. Fogarty used climate surveys in combination with the techniques of multivariate analysis to capture elements of the accident causation process and to test different models of how the components of the system work to conduct a study in which the main aim was to build a model that captures the major sources of variance in maintenance errors. Fogarty argues that these models then can be used to direct interventions aimed at improving safety performance in the maintenance environment.

Although admitting that their study may generalize most to novices because professional air traffic controllers were not available, Murphy, Smith, and Hancock provide a contribution to the basic empirical understanding in the relationship between task demand and the occurrence of errors of omission and of commission through their study conducted in a simulated En-Route air traffic control environment.

Readers who are interested in the human performance issues concerning pilots' use of GPS technology in aviation may find this next article to be of interest. Focusing specifically on the use of cockpit-mounted GPS receivers, Adam, Deaton, Hansrote, and Shaikh performed an investigation with the goal of determining appropriate recommendations pertaining to safe, efficient, and effective interface design and to provide an understanding of the effects of GPS use on pilot behavior and performance.

Using the Piper Warrior instead of a helicopter, the Clarke, Deaton, Villaire, and Shaikh study was an extension of research performed by the Navy, to determine a more effective way to provide information regarding the current state of vital systems or components that could reduce errors and alert pilots to potential malfunctions. Using two prototype interfaces that they developed and tested for a system called Small Aircraft Maintenance Monitoring System (SAMMS), Clarke, Deaton, Villaire, and Shaikh compared the two prototype interfaces, which provided more direct and detailed information about a failure, to the current mechanical fault detection system in the Piper Warrior.

This issue includes a second article for military aviators. Gawron presents her study of the research conducted on the effects of G on humans.

Training Development Reports, Studies, and Papers

This next article may be of interest to those involved in university academic and flight training. One attempt to alleviate the developing pilot shortage in the U.S. airline industry and bridge the gap between the university academic and flight training environment to airline cockpits is presented in the Karp article. He presents a training model that includes flight training focused on airline-type crew procedures and checklists, the use of specific airline flight training devices and motion-based simulators, and cooperative student candidate selection and employment interview agreements between regional airlines and universities.

Today's leaders, as well as future leaders, in aviation may glean some interesting points from the Kutz and Bliss article. They provide a combination of a literature review summary and the findings of a qualitative study of the management and leadership styles and relational competencies exhibited by aviation leaders.

After presenting an extensive background on deregulation, its effects on the airlines, and the changes in flight scheduling, Shank and Sherman relay some of American Airlines' experiences with hub depeaking.

Book Reviews

Our book review for this issue, *Air Rage – The Underestimated Safety Risk* by Angela Dahlberg, is provided by Deak Arch and Mark Sherman. They argue that

the book's format makes it readable and useable. Dahlberg presents available information on air rage, from the situations that provoke the perpetrators to the steps to be taken to contain outbreaks. Recommending the book to all aviation employees, as well as the public in general, Arch and Sherman encourage university-level human factors instructors to utilize the text.

B.S.L.

International Journal of Applied Aviation Studies

Formal Papers

Training Development Papers

Book Review

Formal Papers

Time Series Based Objective Pilot Performance Measures

Nicholas R. Johnson,

Esa M. Rantanen, and Donald A. Talleur

Institute of Aviation, Aviation Human Factors Division
University of Illinois at Urbana-Champaign
Savoy, IL 61874

Abstract

Automated objective pilot performance measures potentially enhance and expand traditional proficiency evaluation methods by an instructor pilot. Quantitative performance data also can be utilized in research and subjected to various analyses to reveal covert patterns in pilots' performance. In spite of a relatively long history of research, routine use of objective measures is still rare. Considering the nature of many flight tasks, it is relevant to investigate the behavior of various flight parameters over time. In this paper, we describe the development of nine specific metrics of pilot performance derived from time series of different flight parameters and examine their descriptive power and sensitivity against data from two pilots with known differences in performance (good and poor), as judged by an expert instructor pilot. Two autocorrelation based metrics and seven Fourier analysis based metrics are evaluated. Initial results showed these metrics to be both sensitive and diagnostic in differentiating between good and poor pilot's performances as determined by the instructor pilot. The findings are consistent with the hypothesis that a skillful pilot will control the aircraft with a greater range of frequencies of input than a less skillful pilot would, making adjustments appropriate to the circumstances whereas a poor pilot appears to make the same adjustment regardless of the actual magnitude of the adjustment needed. The results showed the potential usability of performance metrics derived from time series data and that it is possible to discriminate between good and poor pilot performance using this approach.

Introduction

Objective pilot performance measures are very desirable for a multitude of purposes and for many reasons. Automatic data collection has the potential to enhance and expand traditional proficiency evaluation methods by an instructor pilot by alleviating the time constraints and information overload often associated with direct observation. Furthermore, quantitative performance data can be utilized in research and subjected to various statistical analyses to reveal underlying, covert patterns in pilots' performance. Not surprisingly, objective pilot performance measures derived from flight data recorders (FDRs) or data output from simulators have a relatively long history of research (e.g., Gerlach, 1972; Vreuls et al. 1975; Stave, 1977; De Maio, Bell, & Brunderman, 1985; Benton, Corriveau, & Koonce, 1993). In spite of this, however, it appears that a relatively small number of distinct objective metrics have been utilized in research, and routine use of objective measures are still rare. There are several notable obstacles to application of these measures. For example, Vreuls and Obermayer (1985) noted that the internal processes that drive operator actions are not observable and that few theories of human performance exist to predict what should be measured and the relative importance of each measure. Furthermore, task segmentation is necessary for automated performance measurement, making the process difficult. For maximum utility in training, performance measures need to be available and discernible in real time or as close as possible to the completion of the training session as well (Vreuls & Obermayer, 1985). However, while these problems are undeniable and difficult to overcome (c.f., Rantanen & Talleur, 2001), they are arguably outweighed by the potential benefits of objective measures, making continual research on the latter important.

Objective Pilot Performance Measures in Aviation Research

One of the most common objective metrics is the standard deviation (SD) of selected flight parameters. This metric describes the amount of variability around the mean of any series of values. A small SD in the case of piloting an aircraft will usually be indicative of good performance. For example, Svensson, Angelborg-Thanderz, Sjöberg, and Olsson (1997) examined the effects of information complexity on pilot mental workload and pilot performance in a simulator and found that altitude deviations increased and correction of errors were delayed as a result of increased workload. In addition, Hill and Eddowes (1974) used SD to distinguish between beginning, intermediate, and advanced pilots (based on hours flown) during simulator flights; SD variables produced the highest proportion of statistically significant differences between groups (32%), followed by tracking variables (18%). It is important to note, however, that SD does not provide any information about possible error relative to given criteria.

Root mean square error (RMSE) is a widely used measure of tracking performance (e.g., Scallen, Hancock, & Duley, 1995). It can be used to reduce the tracking performance along a specified parameter value, or criterion (e.g., a given altitude, or VOR radial), in the entire segment of a flight into a single

number. A low number typically indicates good performance. The RMSE is calculated by squaring individual errors (sampled at certain rate), adding them together, dividing this sum by their total number, and then taking a square root of this quantity. The RMSE hence summarizes the overall error. In a study by Reising, Ligget, Solz, and Hartsock (1995) the RMSE measurements were successfully used to reveal pilot performance differences when using two different types of head-up displays (HUDs). Subjective feedback from the pilots corroborated the results. Ververs and Wickens (1996) measured mean absolute errors of altitude, heading, and airspeed, along with reaction time to a stimulus event to investigate the effect of clutter and low lighting on HUD assisted flight in a high fidelity flight simulation environment. Tracking error was used to determine that pilot performance was better in a clear sky condition than a cloudy condition, indicating better extraction of aircraft pitch and roll information from outside the aircraft; either from real horizon or the HUD display. Also Stave (1977) measured performance in a simulated helicopter flight task by RMSE from navigational course and an angular deviation from the Instrument Landing System (ILS) approach.

The RMSE has a number of shortcomings, however. It does not contain information about the direction of deviations or the frequency of deviations from the criterion. The latter is particularly important dimension of tracking performance, as it would allow for detection of high velocity error in tracking while the position error (measured by the RMSE) might be minimized (Wickens & Holland, 2000). To overcome these limitations, additional measures of tracking performance are available.

The number of deviations outside tolerance (ND) is a measure that tallies the occurrences of the aircraft straying outside predetermined tolerances (Reynolds, Purvis, & Marshak, 1990). This is essentially a measure of velocity error in tracking and it complements the RMSE, which contains the error magnitude information. A low number typically indicates good performance. A low value, however, can also be obtained if the pilot makes few aberrations outside the tolerances but stays there for a substantial proportion of the flight segment of flight. The ND measure must hence be considered together with the total time spent outside tolerance in a given segment. The cumulative time the aircraft spends outside a given tolerance provides an indication of tracking performance beyond the RMSE and number of deviations. This measure is computed simply by summing the time the pilot spends outside of a given tolerance and divided by the total time in the segment (i.e., percent time outside tolerance). A small number indicates good performance. Sirevaag et al. (1993) took aircraft control measures from a helicopter simulator in a study investigating the effects of verbal and digital communication loads on pilot performance. The measurement of time above an altitude criterion produced significant differences between experimental task conditions.

Rantanen and Talleur (2001) developed a metric labeled mean time to exceed tolerance (MTE). The MTE is computed from the rate of change between successive data points and the aircraft's position relative to a given tolerance. Based on this information, the measure extrapolates the time the aircraft will remain within the tolerance region, as opposed to the number of deviations and time outside tolerance measures described above. Because this measure could potentially yield very large values, it was truncated at 60 s. Thus, if the pilot was 60 s or more from exceeding tolerance throughout the flight segment, his or her performance was considered good. In subsequent analysis, the MTE on ILS localizer tracking showed a significant difference between pilots who passed an IPC flight and those who failed, by flight instructor evaluation (Rantanen & Talleur, 2001).

Other objective metrics include critical control input, which is defined as a pilot input that changed or led to a change from positive vertical acceleration to negative vertical acceleration (or other flight parameter) or vice versa (De Maio, Bell, & Brunderman. 1985). A non-critical control input did not cause the vertical acceleration to change from positive to negative or vice versa. De Maio, Bell, and Brunderman (1985) hypothesized that "efficient" control would be characterized by a relatively large proportion of critical control inputs indicating that pilots were canceling small errors in altitude frequently. Another metric, "smoothness," was defined as the proportion of critical control inputs from the total number of inputs (critical + noncritical). The critical error rate is the horizontal distance traveled from critical control input to vertical acceleration sign change divided by the time from critical control input to vertical acceleration. This metric was designed to measure the effectiveness of a critical control input; low values for critical error rate would indicate a slow accumulation of error following the pilot control input. De Maio, Bell, & Brunderman (1985) found that that smoothness and critical error rate were affected by flight task difficulty (straight vs. turning, both while level).

A Case for Time Series Analysis

The measures reviewed above can be viewed as static, however. They also average out variations in performance over the course of the flight segment or maneuver being analyzed. Given the dynamic nature of piloting an airplane, it would seem relevant and informative to develop measures that better capture the time-dependence of given flight parameters. Time series analysis (e.g., Bloomfield, 1976; Chatfield, 1975; Gottman, 1981) provides tools that allow for examination of diverse aspects of the time history of data. Two frequently used techniques of time series analysis are autocorrelation and Fourier (or spectral) analysis (Box, Hunter, & Hunter, 1978; Butterfield, 2001). These analysis techniques utilize the time-dependence of the data series to uncover patterns in the time series that would not be brought to light with any of the "static" measures discussed above. That is, latent periodicities and correlations within the time series may be detected using time-series analysis.

In this paper, we describe the development of nine specific metrics of pilot performance derived from time series of various flight parameters and examine their descriptive power and sensitivity against data from two pilots with known differences in performance (good and poor), as judged by an expert instructor pilot. To the best of our knowledge, this represents a novel approach to pilot performance analysis within the general aviation environment.

Method

The metrics described in this paper were developed to supplement existing performance metrics (Rantanen & Talleur, 2001) with analyses that examine underlying patterns in the pilot-generated time series of data. The new metrics utilize spectral (Fourier) and autocorrelation analyses. Two guiding hypotheses were used to develop these metrics: First, there may be a difference in the frequency of observed flight characteristics (based on pilot's control inputs), better pilots exhibiting a larger range of frequencies of aircraft control than less able pilots, who may only control the aircraft with low frequency control inputs. Using Fourier analysis, a time series of data can be decomposed into spectral or frequency components. This decomposition allows an explicit representation of the underlying frequencies occurring in the time series. The second hypothesis is that more skillful pilots will exhibit a better awareness of the airplane's constantly changing state and be able to predict what control inputs will be required to pilot the airplane to the desired state in the future. This may be manifested in the degree of correlation between flight parameter values in a time series. That is, better pilots may exhibit a greater correlation between a previous time point and the present time point than less skilled pilots who may exhibit a greater randomness of control on flight parameter values. By taking the autocorrelation of a time series for a particular observed flight variable, the degree of randomness between successive measurements can be investigated. Derivation of specific metrics from time series data is described next.

Fourier Analysis Metrics

To examine the periodic components of a time series, Fourier analysis was used. Taking the Fourier transform of time series data, Y_k, gives the spectral decomposition:

$$Y_k = \frac{1}{N} \sum_{j=1}^{N} \tilde{Y}_j e^{\frac{2\pi i}{N}(k-1)(j-1)}$$

where the Fourier coefficients $\tilde{Y}_j$ are given by

$$\tilde{Y}_j = \sum_{k=1}^{N} Y_k e^{\frac{-2\pi i}{N}(k-1)(j-1)}$$

and where N is the number of time series data points and $i = \sqrt{-1}$

The original time series is then expressed as a weighted sum over all frequencies contained in the Fourier transform. The weights, $\frac{|\tilde{y}_j|^2}{N}$ represent the contribution a particular frequency makes to the original time series and are termed *power spectral densities* (PSD). The $\frac{|\tilde{y}_j|^2}{N}$ can be plotted against frequency, $f = \frac{j}{N}$ (for a 1Hz sampling rate), in a periodogram.

We hypothesized that a good pilot's time series may contain a greater range of frequencies that contribute significantly to the time series, that is, a greater proportion of components that have a large PSD, compared to a poor pilot's time series. The metrics that were developed with Fourier methods are used to quantify both the range and magnitude of these significant frequency components.

In determining what spectral components of the Fourier decomposition were significant, a critical value v_c was set. Components with PSD greater than v_c were counted and used in the subsequent metrics described below. Setting v_c involves some difficulties, however. Because the data ranges of the time series vary greatly between flight parameters (altitude and airspeed for example) and individual pilots, PSD magnitudes in the Fourier decomposition will also vary greatly between parameters and pilots. Thus setting a single critical value to be used across all pilots' flight parameters will not achieve the desired level of sensitivity. Therefore, a relative v_c was set to a fraction of the mean or maximum value of the spectral components. This approach will also allow for manipulation of v_c in order to find the value that produces maximum sensitivity in distinguishing good and poor pilots.

Seven Fourier-analysis based metrics were developed; (1) mean and (2) standard deviation of the spectral components $\frac{|\tilde{y}_j|^2}{N}$, (3) the number of spectral components that are greater in magnitude than a critical value v_c, (4) the mean and (5) standard deviation of spectral components greater than v_c, and (6) the mean frequency and (7) standard deviation of the frequencies of spectral components with magnitude greater than v_c (see also Table 1).

Table 1

Metrics derived from time series analyses of pilot performance

Metric Code	Metric
AC 1	Slope of the Regression line; quantifies how quickly autocorrelations tend to zero.
AC 2	Least Squares Error; goodness of linear fit to the first 10 autocorrelations measure.
FT 1	Mean of the $\dfrac{\left\lvert \tilde{Y}_j \right\rvert^2}{N}$ the normalized squared magnitude of the spectral components incorporates the magnitude of deviations in the time series.
FT 2	SD of the $\dfrac{\left\lvert \tilde{Y}_j \right\rvert^2}{N}$;
FT 3	Number of $\dfrac{\left\lvert \tilde{Y}_j \right\rvert^2}{N} > v_c$.; number of spectral components greater than a criterion value.
FT 4	Mean of the $\dfrac{\left\lvert \tilde{Y}_j \right\rvert^2}{N} > v_c$; mean magnitude of the spectral components above criterion.
FT 5	SD. of the $\dfrac{\left\lvert \tilde{Y}_j \right\rvert^2}{N} > v_c$; magnitude spread of spectral components in FT 3 above.
FT 6	Mean frequency of the $\dfrac{\left\lvert \tilde{Y}_j \right\rvert^2}{N} > v_c$; mean frequency of spectral components found in FT 3 above.
FT 7	SD of frequencies of the $\dfrac{\left\lvert \tilde{Y}_j \right\rvert^2}{N} > v_c$; the frequency spread of the spectral components found in FT 3

Autocorrelation Metrics

The autocorrelation coefficient (r_h) gives a measure of the correlation between data points Y_k and Y_{k+h} of the time series $Y = \{ Y_1, Y_2, \ldots, Y_N \}$ and is given by:

$$r_h = \frac{\sum_{k=1}^{N-h}(Y_k - \overline{Y})(Y_{k+h} - \overline{Y})}{\sum_{k=1}^{N}(Y_k - \overline{Y})^2}$$

where

$$\overline{Y} = \frac{1}{N}\sum_{k=1}^{N}Y_k$$

is the mean of time series data and $-1 \le r_h \le 1$

A plot of r_h versus lag, h, is termed a *correlogram*; $r_0 = 1$ by definition. The autocorrelation coefficient gives a measure of how well a subsequent measurement can be predicted from a previous value in the time series. Values of r_h close to zero indicate little correlation between data points and values close to –1 or 1 indicate a strong negative or positive correlation respectively between data points. The time series of flight parameter values from a pilot who is aware of the state of the aircraft and can predict its future state may generate autocorrelations that are greater than those from a pilot who is not as aware. Since autocorrelations at large lag h tend towards zero, a useful measure that may be indicative of pilot performance is how quickly the series autocorrelations tend to zero.

We hypothesized that time series of less skillful pilots would produce autocorrelations that decay more quickly to zero than those of more skillful pilots. Consequently, two specific metrics were developed: (1) To quantify the decay of autocorrelation coefficients, the slope of the first 10 autocorrelation coefficients (from lag = 0 to lag = 9) was determined by regression analysis, and (2) the sum of squares error of the fitted regression line was also included as a second autocorrelation based metric.

Data Collection and Analysis

The data were collected from instrument proficiency check (IPC) flights in an aircraft equipped with a data logger that measured airspeed, altitude, vertical speed, heading, pitch, roll, ball deflection, course deflection indications (CDI) and glide slope indications (GS CDI) (see Lendrum et al., 2000; Rantanen & Talleur, 2001). Each flight parameter was sampled at 1 Hz and data stored on an on-board PC. An IPC flight consists of 14 distinct segments, including VHF Omnidirectional Range (VOR) tracking, Instrument Landing System (ILS) approach, non-precision VOR approach, holding pattern, and steep turns. Data from the flights were segmented using a data visualization tool (see Rantanen & Talleur, 2001, for a description of the tool and data preprocessing procedure) before further analyses.

Since this was a proof-of-concept study, data from only two representative pilots were selected for time series analysis. One of the pilots was an expert pilot with good performance during the IPC flight. Data from another pilot was selected based on poor performance during an identical flight, as judged by a flight instructor. Time series analysis was carried out in *MATLAB* version 6.5.1. The analysis generated an array of 1134 metrics (14 segments and 9 flight parameters, each with 9 metrics). This number contained several metrics that were not useful (for example, CDI or GS CDI measurements are not valid on flight segments that did not use these instrument indications) and hence were not analyzed further. Detailed results from four particular flight parameters in one flight segment are presented next.

Results

We analyzed time series of altitude, vertical speed (VS), airspeed (IAS), and Course Deviation Indicator (CDI) deflection from a VOR approach segment of an IPC flight. This segment involved tracking of a VOR radial inbound to final approach fix (FAF) and it was chosen for initial analysis since it was an essentially unaccelerated straight-and-level tracking maneuver. Although several flight variables recorded on this segment provided quantifiable information about the proficiency of the pilot's flying skills, vertical speed (from the vertical speed indicator, VSI) in particular was a good indicator of the level of difficulty the pilot had holding the assigned altitude. Because the VSI is quite responsive to small changes in aircraft pitch, it provides a more sensitive measurement of short-term motion than the altimeter. While a pilot may keep the airplane within the specified altitude tolerances during the flight segment, the vertical speed may oscillate rapidly and with large changes in magnitude. This control characteristic is indicative of poor piloting technique that altitude analysis alone may not diagnose. Qualitative graphical analysis showed that VS provided the best distinction between good and poor pilot performances of the four flight parameters used in analysis of the VOR tracking flight segment.

The raw time series data of these parameters from this segment are shown in Figure 1. The critical value used to count a spectral component of the Fourier series was set at one eighth the mean value of the series.
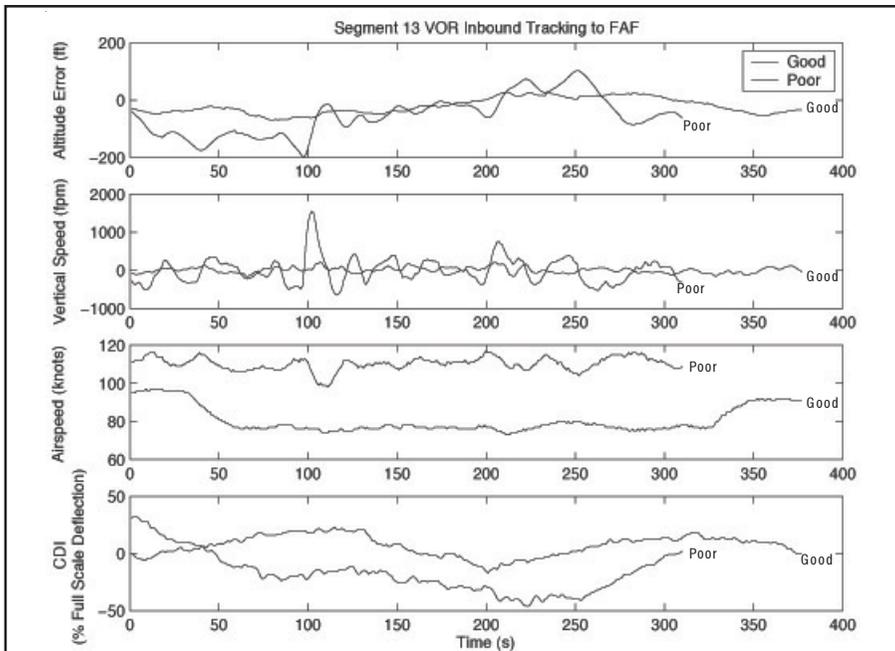
Figure 1. Raw time series data from good and poor pilot (as judged by expert flight instructor) from four flight parameters (altitude, vertical speed, airspeed, and course deviation indicator) from a VOR approach segment of an IPC flight.

Data Visualization and Qualitative Analysis

The data were first plotted in correlograms and periodograms for visualization and qualitative analysis. The differences between good and poor pilot performance can be clearly seen in a correlogram of vertical speed data from a level flight segment during a VOR approach (Figure 2). The good pilot's autocorrelation coefficient decays much slower versus lag than the poor pilot's. The better pilot makes smaller adjustments more often and of a more consistent magnitude than the poor pilot who may be unsure about exactly what action or correction is required. We call the latter behavior a mechanical application of the controls whereas the better pilot uses steady pressures on the controls and "eases" the plane back to the criterion. This is why in a time series there is higher correlation between movements of the VSI in one time period to the next for a good pilot. The poor pilot makes fast and jerky movements and may make multiple movements with different accelerations.
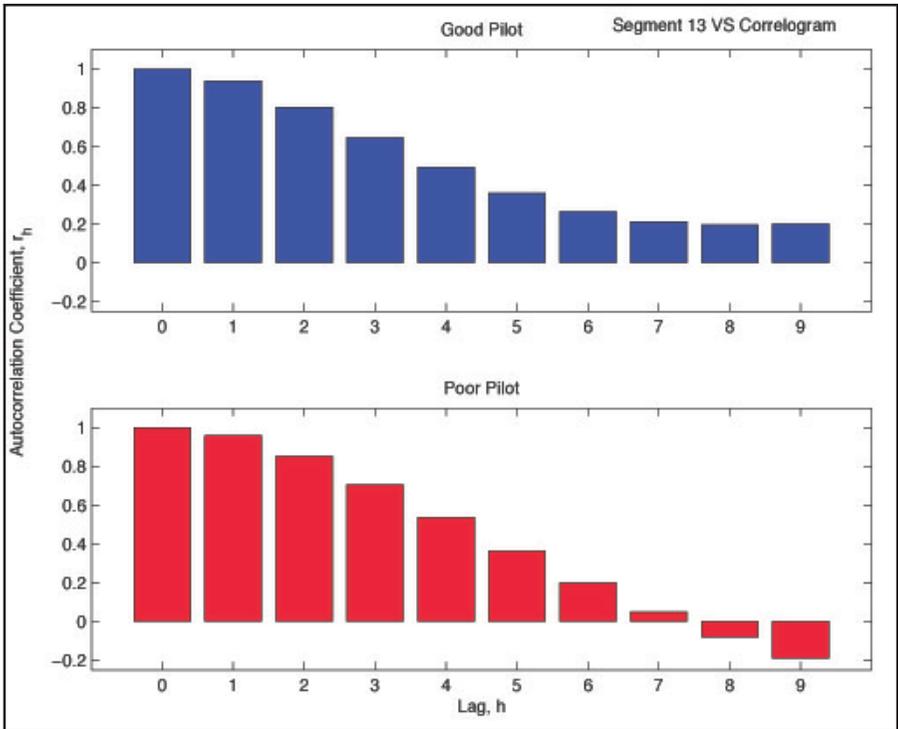
Figure 2.Correlograms of good and poor pilots' vertical speed from the same segment (c.f., the second time series plot in Figure 1). Note that the good pilot's correlation coefficient decays from 1 much slower than the poor pilots does (i.e., it has a shallower negative slope).

The periodograms from the same flight segment and parameter reveal even more substantial differences between the two pilots (see Figure 3). The poor pilot's vertical speed periodogram is characterized by a number of high magni-

tude spectral components at frequencies less than approximately 0.05Hz. This is contrasted by the more even spread of spectral components out to 0.1Hz seen in the good pilot's periodogram. The contrast is consistent with our hypothesis that a good pilot will control the aircraft with a greater range of significant frequencies of control input than a poor pilot would, who will control the aircraft with predominantly low frequency inputs. While the poor pilot's periodogram contains some spectral structure beyond 0.05Hz, it is greatly less significant that the components at frequencies below 0.05Hz. Note that the two periodograms have significantly different y-axis scales. Plotting the good pilot periodogram on the same y-axis scale as the poor pilot periodogram would not allow the detail of the good pilot's distribution of spectral components to be seen. The Fourier metrics were created to quantify the observed qualitative differences in the distributions of spectral components of the good and poor pilot periodograms.
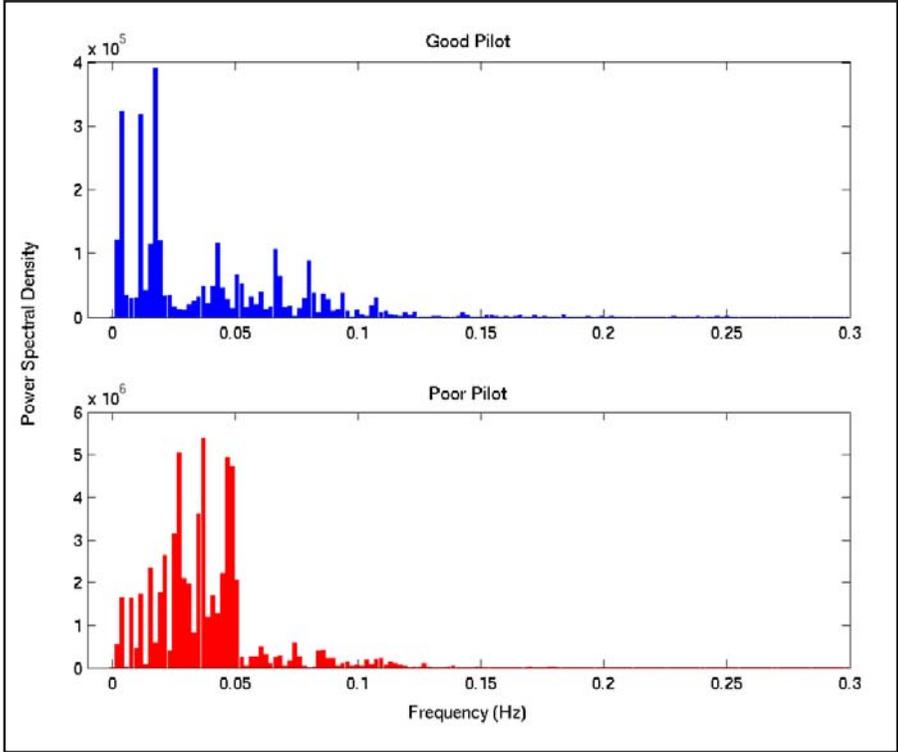
Figure 3. Periodograms of the same data depicted in the second time series plot in Figure 1 and in Figure 2. Note the smaller magnitude (by an order of magnitude; the y-axes have different scales) and greater number of spectral components in the good pilot's time series than the poor pilot's time series.

Time Series Based Metrics

The nine time series metrics bases on autocorrelation and Fourier analysis

are shown in Table 2 comparing the good and poor pilots' performance in altitude, vertical speed, airspeed and CDI deflection control in the aforementioned flight segment. Differences between the good and poor pilot can be seen consistently across all flight parameters from the autocorrelation-based regression slope metric (AC 1). The slopes are consistently shallower for the good pilot than for the poor pilot. The poor pilot's regression slope was between 43% (Vertical Speed) and 50% (Airspeed) greater than the good pilot's was. This indicates a higher degree of predictability between subsequent time series values for the good pilot than the poor pilot. All values for the R^2 (AC 2) were close to zero, attesting to a good fit of the linear regression line.

Table 2

| Metric	Altitude		VerticalSpeed		Airspeed		CDI	
	Good	Poor	Good	Poor	Good	Poor	Good	Poor
AC 1	-0.0079	-0.0219	-0.1007	-0.1436	-0.0168	-0.0847	-0.0123	-0.0189
AC 2	0.0001	0.0013	0.0569	0.0158	0.0002	0.0117	0.0001	0.0000
FT 1	1.768×10^7	1.729×10^7	1.168×10^4	2.379×10^5	1.301×10^4	2.425×10^4	2.568×10^2	1.204×10^3
FT 2	2.136×10^8	1.874×10^8	4.141×10^4	7.919×10^5	1.552×10^5	2.615×10^5	2.406×10^3	1.119×10^4
FT 3	14	17	80	64	20	17	25	25
FT 4	3.217×10^8	2.594×10^8	3.694×10^4	9.458×10^5	1.660×10^5	3.637×10^5	2.619×10^3	1.228×10^4
FT 5	8.906×10^8	7.036×10^8	6.796×10^4	1.368×10^6	5.457×10^5	9.814×10^5	7.443×10^3	3.456×10^4
FT 6	0.0156	0.0200	0.0879	0.0647	0.0226	0.0201	0.0263	0.0259
FT 7	0.0124	0.0143	0.0598	0.0380	0.0150	0.0145	0.0183	0.0192

The range of Fourier metrics also differentiates between the good and poor pilots, although some inconsistencies appear between flight parameters. The mean (FT 1) and standard deviation (FT 2) of the normalized squared magnitude of the spectral components clearly separate the good and poor performances in vertical speed, airspeed, and CDI deflection control. In all three cases, there is an order of magnitude difference between the good and poor cases. This difference reflects the greater variation in time series values that can be seen in Figure 1 for these flight parameters for the poor pilot compared to the good pilot. The metrics from the altitude data do not separate the good pilot from the poor pilot (FT 1: 1.7678 x 10^7 and 1.7295 x 10^7 respectively and FT 2: 2.1361 $x 10^8$ and 1.8738x x 10^8 respectively). The lack of sensitivity seems to derive from the almost ideal performance of the good pilot in holding a constant altitude over the approximately 6-minute segment. Because the good pilot's time series is very close to a flat line, or alternatively, had low amplitude and low frequency oscillation, the Fourier spectrum of the time series is made up predominately of zero and low frequency components. Hence, values for FT 1 and 2 that are comparable to those attained from the poor pilot.

The remaining Fourier metrics are based on the number of frequency components that are greater than the set critical value (v_c = one eighth of mean spectral components). The number of spectral components greater than v_c (FT 3) from the vertical speed and airspeed data show a difference between good and poor performance, while for the CDI data, FT 3 is equal for the good and poor pilots.

It should be noted that the number of high magnitude spectral components that contribute to the times series will change as a function of the critical value. As v_c is lowered, a larger number of smaller magnitude spectral components will be included in metric FT 3.

The mean (FT 4) and standard deviation (FT 5) of magnitude of spectral components greater than v_c effectively differentiate good and poor pilot performance for the vertical speed, airspeed, and CDI flight parameters. For example, the good pilot's mean and standard deviation of the high magnitude spectral components (FT 4: 3.6941×10^4 and FT 5: 6.7957×10^4 respectively) are an order of magnitude smaller than the corresponding poor pilot values (FT 4: 9.4582×10^5 and FT 5: 1.3677×10^6). This indicates that the high magnitude spectral components that contribute most significantly to the time series are lower and less spread in magnitude for the good pilot than for the poor pilot. Again, this is consistent with the hypothesis that a skillful pilot will control the aircraft with a range of input frequencies, and not a limited number of low frequency inputs, which would lead to Fourier spectrums containing a very limited number of high magnitude components.

The mean (FT 6) and standard deviation (FT 7) of the frequency of spectral components also provide a clear separation between good pilot and poor pilot performance for the vertical speed and CDI control. The contrast is greatest for the vertical speed metrics, where the mean frequency of the high magnitude spectral components (FT 6: 0.0879 Hz) and their standard deviation (FT 7: 0.0598 Hz) for the good pilot are greater than those for the poor pilot (FT 6: 0.0647 Hz and FT 7: 0.0380 Hz). This shows the good pilot has both a larger mean frequency and a greater range of high magnitude spectral components than the poor pilot does, as our hypothesis of skilled pilot performance predicted. For the CDI, however, these metrics do not offer the same ability to discriminate between good and poor performance.

The Fourier metrics for altitude show little difference between good and poor pilot performance, despite the time series being qualitatively different (Figure 1). As discussed above, the good pilot showed so little deviation from the desired altitude over the course of the flight segment that the Fourier spectrum derived from the time series was made up of predominantly low frequency, high magnitude components. Such a spectrum will show little quantitative differences from a poor pilot's spectrum. This may limit the use of such slowly varying time series like altitude in generating effective metrics based on Fourier analysis.

Summary and Discussion

The research reported here represents a proof-of-concept study of novel time series based measures of pilot performance. By our initial analyses, these metrics appear both sensitive and diagnostic in differentiating between good and poor piloting performances as rated by an independent instructor pilot. The

autocorrelation based regression line slope showed that a good pilot generates autocorrelation values that decay from 1 more slowly than the poor pilot's does. The sum of squares regression errors were close to zero indicating a high quality of fit of the regression line to the first 10 data points. Differences between the good and poor pilot can also be seen consistently in the Fourier-based metrics. The larger scale of variation and range of vertical speed of a poor pilot from the original time series data was also well represented by the Fourier metrics, which clearly showed that the good pilot had a higher mean frequency of significant spectral components and a larger number of spectral components that contribute significantly to the time series. These findings are consistent with the hypothesis that more skillful pilots control the aircraft with a greater range of frequencies of input than less skillful pilots. This indicates that the skilled pilot makes an adjustment appropriate to the circumstances whereas the poor pilot appears to make the same adjustment regardless of the actual magnitude of the adjustment needed. It may be argued that the idealized "poor" pilot has a single mode of operation (a coarse yank on the controls once he or she has realized a flight parameter is not where it should be) whereas the good pilot will be constantly aware of what control input is required to get the aircraft in the desired state (i.e., AC 1 metric) and be able to perform such inputs with differing magnitude and frequency (FT metrics).

It should be noted, however, that the number of components changes as the critical value changes. The lower the critical value, the number of spectral components that will be counted and used in metrics FT 4-7 will increase. We have observed this in a limited number of segments. This, in turn, will alter the ratio of spectral components counted (and considered significant) between the good and poor pilots. Further analysis is needed to determine how a critical value that produces the greatest separation between good and poor pilots could be set and how it may affect the Fourier metrics. It must also be acknowledged that some of these metrics may not effectively differentiate good and poor performance on other segments or flight parameters. Altitude varies slowly compared to vertical speed and this may limit the use of altitude data for these specific metrics based on time series analysis. Further work will investigate which metrics are most effective in separating good pilots from poor pilots by various statistical analyses (e.g., factor- or principal components analysis) on data from a large number of pilots with different skill levels and demonstrated performance. The prospect of developing an algorithm for setting the criterion used by metrics FT 3–FT 7 dynamically to maximize these metrics' power in separating well and poorly performing pilots is certainly worthy of further investigation.

Other caveats include the issue of segmenting, as improper segmenting of a flight maneuver can lead to spurious results. In addition, the overall variability both within and between pilots across all flight segments and parameters must be addressed, as well as the criteria (e.g., instructor pilot judgment) against which objective metrics are to be validated. However, the results reported here clearly indicate the feasibility of deriving objective performance metrics from

autocorrelation and Fourier analysis of time series data of flight parameters and that it is possible to discriminate between good and poor pilots' performance using this approach.

Conclusion

The metrics that were developed in this study will ultimately be compared to instructor pilot ratings of the pilot being examined in the IPC flight. To be of use in assisting subjective judgments, these metrics must agree with ratings given by the instructor pilot. In other words, if objective metrics are to be used for pass/fail judgments, they must capture the same information as a human observer (i.e., an instructor pilot). Furthermore, valid criteria must be established for each metric. Other requirements for objective metrics include sensitivity in illustrating differences between pilots and their diagnosticity, that is, whether the metrics are understandable by instructor pilots reviewing and interpreting them.

The utility of the proposed metrics in evaluating pilot performance must be validated by statistical analysis from a large number of pilots currently involved in experimental research. This analysis will include data from all flight parameters across all relevant flight segments from approximately 75 pilot subjects. Factor- or principal components analysis will then be used to determine which metrics best agree with the subjective instructor pilot evaluations. It should be noted, however, that particular metrics may not agree with instructor pilot evaluations because of the limitations of subjective evaluations, as discussed earlier. The results presented here nevertheless show that differentiation between good and poor pilot performance based on the described metrics is possible.

Acknowledgements and Disclaimer

References

Benton, C. J., Corriveau, P., & Koonce, J. M. (1993). *Concept development and design of a semi-automated flight evaluation system (SAFES).* (AL/HR-TR-1993-0124). Brooks Air Force Base: Armstrong Laboratory, Human Resources Directorate.

Bloomfield, P. (1976). *Fourier analysis of time series: An introduction.* New York: John Wiley & Sons.

Box, G. E. P., Hunter, W. G, & Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design data analysis, and model building*. New York: John Wiley & Sons.

Butterfield, A. J. (2001). *Data mining of NASA Boeing 737 flight data: frequency analysis of in-flight recorded data.* (NASA/CR-2001-210641). Hampton, VA: NASA Langley Research Center.

Chatfield, C. (1975). *The analysis of time series: Theory and practice.* London: Chapman and Hall, Ltd.

De Maio, J., Bell, H. H., & Brunderman, J. (1985). *Pilot-oriented performance measurement*. (AFHRL-TP-85-18). Brooks Air Force Base: Air Force Human Resources Laboratory.

Gerlach, V.S. (1972). *Model and procedures for an objective maneuver analysis* (Tech. Rep. No. 21201). Arizona State University: Instructional Resources Laboratory.

Gottman, J. M. (1981). *Time series analysis: A comprehensive introduction to social scientists.* Cambridge: Cambridge University Press.

Hill, J. W., & Eddowes, E. E. (1974). Further development of automated GAT-1 performance measures. Final Report 73-72. Air Force Human Resources Library. Abstracted in *JSAS Catalog of Selected Documents in Psychology, 5*, 222; 1975.

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

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

Reising J. M., Ligget, K. K., Solz, T. J., & Hartsock, D. C. (1995). A comparison of two head up display formats used to fly curved instrument approaches. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Reynolds, M. C., Purvis, B. D., & Marshak, W P. (1990). A demonstration/ evaluation of B-1B flight director computer control laws: A pilot performance study. *Proceedings of the IEEE National Aerospace and Electronics Conference* (pp. 490-494). Piscataway, NJ: IEEE

Scallen, S. F., Hancock, P.A., & Duley, J A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics, 26* (6), 397-403.

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics, 36* (9), 1121-1140.

Stave, A. M. (1977). The effects of cockpit environment on long term pilot performance. *Human Factors, 5*, 503-514.

Svensson, E., Angelborg-Thanderz, M., Sjoberg, L., & Olsson, S. (1997) Information complexity –mental workload and performance in combat aircraft. *Ergonomics, 40* (3), 362-380.

Ververs, P. M., & Wickens, C. D. (1996). The effect of clutter and lowlighting symbology on pilot performance with head up displays. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* . Santa Monica, CA: Human Factors and Ergonomics Society.

Vreuls, D., & Obermayer, R. W. (1985). Human-system performance measurement in training simulators. *Human Factors, 27* (3), 241-250.

Vreuls, D., Wooldridge, A. L., Obermayer, R. W., Johnson, R. M., Norman, D. A., & Goldstein, I. (1975). *Development and evaluation of trainee performance measures in an automated instrument flight maneuvers trainer* . (Report NAVTRAEQUIPCEN 74-C-0063-1). Orlando, FL: Human Factors Laboratory, Naval Training Equipment Center.

Wickens, C. D., & Holland, J. G. (2000). *Engineering Psychology and Human Performance* , 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Enhancing the Policymaking Process in Air Transportation:

Application of the Policy Research Construct

Brent D. Bowen
The University of Nebraska at Omaha
6001 Dodge Street
Allwine Hall 422
Omaha, NE 68182.0508
Voice 402.554.3424
Fax 402.554.3781
bbowen@mail.unomaha.edu

and

Chien-tsung Lu
Central Missouri State University
Department of Aviation
Warrensburg, MO 64093
Voice 660-543-8092
ctlu@cmsu1.cmsu.edu

Abstract

While public-policy research has recently undergone rapid development, a longtime barrier between policy analysis and policy evaluation has reduced the accuracy and effectiveness of traditional policymaking. Policy research in the postmodern era utilizes dynamic interaction and reciprocating communication to bring together policy recipients, local communities, and policy researchers for better decision-makings. Citizen participation involving all policy enthusiasts in commenting on policy proposals not only can weed out irrelevant issues, it can also efficiently collect valuable public input. By utilizing both policy analysis and evaluation simultaneously, researchers can serve the public's best interests. Policymaking in aviation is no exception to this observation, as the

ongoing Small Air Transportation System (SATS) project of the National Aeronautics and Space Administration (NASA) and its research partners can attest. This article seeks to address the aforementioned regulatory challenges by proposing a new policymaking tool, namely policy research construct (PRC). The PRC is a proposed research framework containing the merits of both policy analysis and evaluation and is an attempt to bridge the gap between the policy dyad. Ultimately, this paper hopes to provide a malleable and comprehensive tool for current and future aviation policymakers.

Introduction

Traditional aviation policy analysis has been primarily based on empirical data. But without an in-depth exploration of local problems, needs, and context, such analysis could be less useful than it should be in informing policy decisions (Bowersox, 2000). For example, the needs of minority groups in our society may not be identical to those of majority groups for certain policy questions such as clean air policies in southern California (Bae, 1997). And while metropolitan cities take services such as air transportation for granted, remote communities should be able to share the same advantages (Bowen & Hansen, 2000; Bowen & Tarry, 2001; Rives & Serow, 1984; Scottish Council for Voluntary Organizations, 2001). To this end, the role and research methods of policy designers and legislators should be enhanced to include these contexts (Centre for Research and Learning in Regional Australia, 2000; Fox & Miller, 1996).

The Airline Deregulation Act of 1978 led to commercial aviation's implementation of hub-and-spoke logic and infrastructure for the highest possible share of passengers. However, the hub-and-spoke system has also resulted in current gridlock around hub airports both in the air and on the ground. This situation has triggered eminent policy concerns related to unforeseen challenges in navigation, training, and restructurization for delay reduction (Blake, 2000; North, 1999). In the same vein, the launch of the Small Aircraft Transportation System (SATS) will eventually face similar questions, and policymakers should proactively prepare for the upcoming challenges. While a broader utilization of general aviation airports in the U.S. can be envisioned due to the development of SATS, it also imposes crucial policy issues for policymakers to solve. Issues related to airport infrastructure, airspace usage, technology, licensing, training, noise mitigation, and security must be treated carefully before the implementation of SATS. While a linear policymaking process is controversial (Lu, 2003), a more effective policymaking framework would aid policymakers for both SATS and other commercial aviation.

Background

Many social researchers have asserted that when policy researchers embrace only traditional scientific discourse, it is difficult for policymakers to have a functional channel with which to accurately reflect local needs (Bernstein,

1983; Rorty, 1982). Robson (1993) highlighted the essentials of effective communication and exchange between policymakers and local community members in the postmodern era. Policy research, according to Robson (1993), is an invitation to external participation as well as the vehicle for those concerned with determining policy. Among postmodern public administration scholars, Fox and Miller (1996), Smith (1998), and White (1999) argued and stated that scientific enlightenment's promises of universal truth and unified meaning of a specific phenomenon are, to some extent, inadequate. Indeed, policy action research focuses on participation in lieu of theoretical concepts (White, 1999). In today's citizen-participation society, retrieving opinions and suggestions from local residents can be the key to successful policymaking, guiding policymakers to generate the most effective public goods (Box, 1998; Hakim, 2000; Tritter, 1995). Phil Nyden and associated researchers also remarked that policy research should be "an analysis of community needs and potential ways of addressing those needs" (Nyden, Figert, Shibley, & Burrows, 1997, p. 8). Not sporadically, policymakers have habitualized themselves either in espousing discourses of policy analysis (i.e., documents or personal knowledge) or policy-evaluation (i.e., numbers or statistical results) (Smith, 1998). But when their decision-making proves ineffective, the reengineering of previous policies as well as dealing with the consequences of those policies are expensive and time-consuming (Richardson, 2001).

In the aviation field, as Bowen and Hansen (2000) initially outlined for SATS, policy research has applied research tools in policy decisions; but such research itself was yet to be recognized as a tool. Schaaf (2001) adopted Bowen and Hansen's policy construct to a research project on air rage—successfully identifying the legislative gaps, the causal linkages to air rage, and possible solutions for policy changes. Bowen and Tarry then introduced their enhanced policy research construct to the SATS project in 2001. This revised construct was an attempt to accommodate participation from local communities in policymaking. It allowed local collaborative actions to engage in continuous evaluation and feedback (through reports, updates, and revisions) within a democratic and constantly changing developmental environment (Bowen & Tarry, 2001). Accordingly, policy evaluation measures performance and satisfaction must be judged by the government as well as those who will be impacted by new policies. This study provides another in-depth explanation seeking the maximum interest for the aviation industry.

Review of Literatures

In this study, five consistently cited articles and books were reviewed. The critiques of the readings formed the basis of policy research construct.

Majchrzak (1984) started her policy research book, *Methods for Policy Research*, with an elaborate definition of policy research: "the process of conducting research on, or analysis of, a fundamental social problem in order to provide

policymakers with pragmatic, action-oriented recommendations for alleviating the problem" (p. 12). From this research, an array of operations can be simultaneously confirmed. Majchrzak's operational process of policy research is indifferent to that of conventional social science. She stated that policy research should begin with formulating research concepts and then continue with articulating research questions, designing data collection, identifying the analysis method, developing malleable functions, and communicating to lead policymakers. In addition, because provisional policy recommendations may ultimately face challenges in implementation, the range of data collection should not be limited to a narrow domain. The data input should involve all related researchers and key personnel. Thus, research questions should be open in nature. Policy research must be a multi-dimensional, empirico-inductive, malleability-oriented, reciprocating, and communicating process (Majchrzak, 1984).

Majchrzak's work particularly discussed the mechanisms of data collection and technical analysis. If we tend to regard policy research as a complete research structure, data collection and analysis should be an essential component underpinning the entire research skeleton. Majchrzak (1984) proposed several data-collecting tools for beginning researchers: interviews, case research, qualitative research, field research, focused meta-analysis, secondary analysis, surveys, field experiments, and cost-effect analyses. Because policy research should be an empirico-inductive process, the data-collection tool does not need to be precisely framed. In fact, policy research often adopts a combination of several methodologies from which multifaceted themes can emerge (Majchrzak, 1984). In order to create the highest verification and trustworthiness, Majchrzak strongly suggested the usage of statistical analysis and tests such as t-test and Pearson's coefficient correlation analysis.

Hutjes (1991) added historical perspective to applicable policy research strategies. Policy research, Hutjes argued, should be a dynamic tool embedded with dual characteristics: generating and testing. She stated that any research project's tentative findings or policy suggestions should be re-inputted into the policy-decision machine to further modify consequences and possible policy decisions. Reciprocal procedures would eventually move the policy closer to mirroring true public desires. In particular, Hutjes argued, policy research should fulfill four mandates: 1) focus on the design and interpretation phase of research; 2) interact with diverse postulates and crystallize the diagnostic process; 3) conduct in-depth investigations of an unsolved problem; and 4) select the best channel for implementing results. There is also no separate application of qualitative methods in Hutjes' policy research construct (1991). In addition, she recommended that policy researchers apply quantitative instruments.

Like typical policy researchers, Carol Weiss (1991) also addressed why tentative results of policy research should reenter the policy milieu for final policy decisions. She argued that policy research has developed from traditional procedures into a dynamic model that allows researchers to closely interact with

research participants and gain the highest authenticity. Participants' multi-dimensional feedback is genuine, from which researchers could validate collected data. Based on the scenario of self-interpretation of data, Weiss also addressed three disadvantages of using self-interpretation policy research. First, she stated that policy researchers' interaction with partisans may add purposive interests into decision-making. Furthermore, Weiss argued that research problems might be only partially explained to participants, which may generate subjective answers. Finally, even if the research applies a series of rigorous processing phases, a weak interpreter could easily distort the meaning of collected data.

In his article *Organizational Context, Sponsorship and Policy Research Output*, James M. Rogers (1994) asserted that the resulting process of policy research is constantly affected by external input or influences such as politicians, special-interest groups, lobbyists, and appointed organization leaders. Rogers' statement has echoed Weiss' concern about external influence in research for policymaking. Rogers (1994) pointed out the pitfalls of sponsor-type policy research, arguing unperceived biases could ultimately lead to questions about whether the research is applicable only in a special policy environment or to interpersonal relationships. As Rogers put it, the ongoing existence of intersubjective/interpersonal and organizational problems "stemming from the relations between researchers and clientele" affect policy adoption. It means that sponsor-type research may present policy alternatives that solely reflect "the interests of those who sponsor them" (p. 3). Rogers analyzed the United States Political Science Documents database (USPSD) and revealed two major findings. First, organizational context does influence policy research. Second, the type of sponsorship is important to research performance. Nevertheless, how to counteract the bold influence of sponsors in policy research remained unformulated by the time Rogers published his project in 1994.

Haas and Springer (1998) tried to give an in-depth explanation of policy research in their book, *Applied Policy Research: Concepts and Cases*. They thoroughly addressed their understanding of policy research from introduction, strategies, research design, and possible processing models to the description of applied policy research by case-study format. In the strategies section of policy research, they presented three different linear approaches to policy decisions: policy analysis, program evaluation, and statistical analysis. Haas and Springer argued that policy researchers should select their research strategy based on the purpose of their study, which would allow identification of a proper data collection instrument. Haas and Springer outlined a straightforward operational procedure for conducting policy research in the following steps:

Step 1: Define the policy problem and information needs.
Step 2: Compile issues being addressed, select the appropriate research design, and form operational steps.
Step 3: Collect data specifically for research goals and analyze data.
Step 4: Apply results of data analysis to the problems and report to policy makers and clients (Haas & Springer, 1998, p. 70).

Pursuing the Generality of a Policy Research Methodology

Majchrzak's study did not sufficiently address technical analysis because several of the data collecting tools she introduces (e.g., field research) are unable to collect useful quantitative data for possible "empirico-inductive process" or statistical tests. The process of validating these qualitative tools remains unclear and should be the object of further research. Conversely, policy research should still make tentative recommendations based on the results of data analysis through either quantitative or qualitative vehicles. As stated by Majchrzak, a prediction test must be done in order to enhance the possibility of success and to persuade policy implementation before workable recommendations are forwarded to leading/powerful policymakers. Following probability examinations, these tentative recommendations could be revised, discarded, or accepted by policymakers. Majchrzak argued that such a forecasting test could be done by classical regression analysis or the Bayesian possibility model. However, Majchrzak's proposal of statistical exams becomes irrelevant when a policy researcher can only collect qualitative data. Paradoxically, she did not provide any evaluation of policymaking by qualitative tools, such as focus group, ethnographic studies, field observations, and in-depth personal interviews.

Likewise, Hutjes' research perspective guides us to recognize the merits of quantitative-qualitative approaches regarding the future operation of policy research methodology. Her reciprocity aspect of policy analysis and evaluation is plausible. Unfortunately, Hutjes did not explain her criterion of "satisfaction measurement" for the public, nor did she clarify the best time for policy adoption. Moreover, the article does not stress the significance of policy evaluation. It is risky to evaluate a policy's performance after its implementation.

Rogers' argument (1994) regarding research sponsorship and its direct influence on policy decision triggered a search for unbiased policymaking. As a matter of fact, Rogers raised the concern that a prevailing actor may dominate policy decisions and damage the authenticity of policymaking. However, while Rogers criticized dominating actors in policymaking, he failed to propose solutions to the problem.

Weiss' arguments (1991) identified two major problems for public policy research: omitting reciprocal and discursive data contribution and public auditing in policymaking. If policy research only contains public participation/voice without a reciprocal process in data collection, its findings are less useful. Furthermore, without a close and constant inspection from the participants/the public, even a well-trained policy researcher may tend to insert personal bias or interpret data based on personal perception. For this reason, traditional scientific researchers tightly embrace only statistical analyses to avoid such pitfalls. Although, Weiss did criticize the traditional policy research, she did not provide a workable model for policy researchers.

Haas and Springer's (1998) policy research model is disputable because there is no re-input or evaluation phase in their policy research procedures. The policy research Haas and Springer tried to promote is a straight-line/one-way process—the typical drawback of current policymaking and research via traditional scientific methodology. Haas and Springer gave several policy research examples through case studies that limit the significance and applicability of the research.

<center>Introduction of Policy Research Construct</center>

The above critiques of traditional policymaking have raised research sensitivity. In particular, the authors recognize that policymakers need to hear the voice of related individuals and assess possible results of implementation in the hope of making better decisions. In this study, a new policy research construct, whose goal is to promote aviation policymaking that better meets public needs, is being proposed. The operational flowchart (see Figure 1) indicates the genre of this policy research construct. The attempt is to consolidate research provisions and connect the separate usage of policy analysis and policy evaluation. The construct has three research phases (policy reviews, policy research, and policy actions) containing seven research steps, helping to ground and foster this proposed research framework.
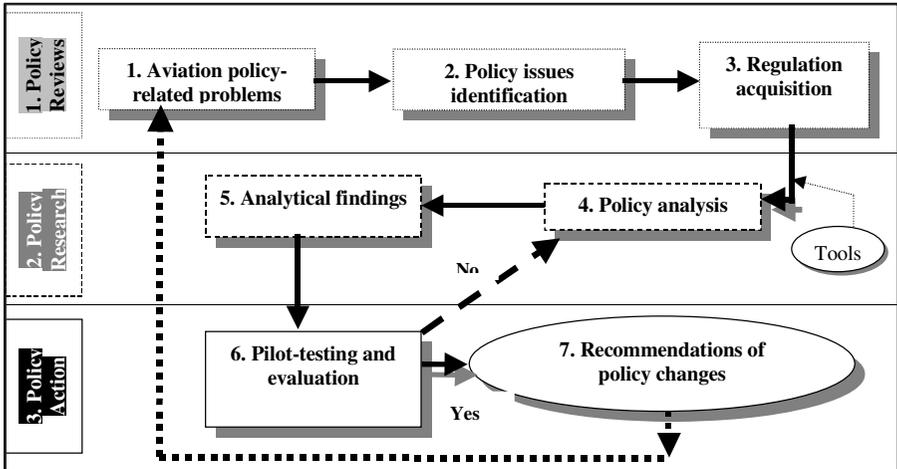
Figure 1. Flowchart for the aviation Policy Research Construct (PRC)

Phase One – Policy Reviews

The research phase of policy reviews differs from traditional policy analysis in that it equally weighs the public voice. This policy-review phase will guide researchers to revisit existing aviation-related policies after encountering a new policy challenge (Steps 1 and 2). As Jenkins (1978) and Walker (1993) stated in terms of conventional policy analysis, policy researchers should revisit existing policy and locate problems such as inappropriateness and obsolescence. The

existing laws and policies related to a specific issue can be retrieved from electronic libraries such as Congressional Universe and Lexis-Nexis. The review of policy helps researchers to identify the factors creating problems and those that need to be readdressed. Accordingly, aviation policy researchers should first review and analyze existing aviation policies related to newly encountered aviation problems and subsequently seek the strategy necessary to cope with deficient policy. Jenkins' book *Policy Analysis: A Political and Organizational Perspective* (1978) provided a linear process of policy analysis from initiation, data collection, consideration, and alternation to policy outputs. Jenkin's research was one of the studies concerning policy review and Phase 1 of the PRC will similarly see the advantages of traditional policy analysis.

The construct's process of data collection/acquisition (Step 3) includes a review of literature primarily focusing on up-to-date Federal Aviation Regulations (FARs) and documents from the Government Printing Office (GPO). Secondary data analysis, if needed, would be performed through analytical tools such as content analysis or historical research.

Phase Two – Policy Research

The proposed policy research construct for aviation policy advocates the retrieval of opinions of the FAA for a real-time reflection and social information update during research Phase 2 (Steps 4 and 5). In addition, researchers should not lose their policy focus as they incorporate public participation. Data collection tools include, but are not limited to, surveys, personal interviews, Delphi technique using reciprocating interactive procedures, public hearings, symposiums, focus groups, and panel studies. Researchers must constantly remind themselves of three critical questions during this phase: (1) Are the policies inadequate or outdated? (2) What would be the consequence without further revisions of related policies? (3) What would be the possible impact if we propose a new policy? Certainly, policy researchers' self-awareness should not be constrained at any point in the policy process.

In addition, policy analysis (step 4 and 5) should simultaneously focus on data analysis via mathematical tools (such as Niskanen's [1998] policy analysis of welfare and the culture of poverty) and data coding (such as Haas and Springer's [1998] housing policy study) to formulate analytical findings, contingent provisions, and tentative postulates. The grounded policy-change results (affiliated with policy-change recommendation) could be justified not only by simulation (Majchrzak, 1984) and economic analysis, but should also be debated by the affected personnel and groups (Bernstein, 1983; Fox & Miller, 1996; Hakim, 2000; Nyden, Figert, Shibley, & Burrows, 1997; Robson, 1993; Rorty, 1982).

Phase Three – Policy Actions

In the policy action phase, pilot-testing (Step 6) and policy recommendations changes (Step 7) obtained from Phase 2, should be evaluated in terms of the

possible impact to future air industry. If recommended policy changes will substantially affect aviation actors (such as local communities, airports, manufacturers, government, and related policy enthusiasts), conducting brainstorming seminars is always beneficial to decision making and policy formation.

Policymakers have two alternatives after the public action phase. First, researchers could carry the recommendation back to upper phases (Step 6) for another in-depth data analysis, provided the draft recommendation could not obtain congruence from participants. Through the recursive and reciprocal process and discussion (following consequent procedures through Steps 4, 5, 6, and 7), the final and most appropriate agreement can be approached upholding policy mandates, revision, or implementation. Second, should the final conversational congruency or alike be obtained, the policy researchers could proceed with what should be implemented in policy mandate or generation (Step 7).

Using PRC Concept— An Example of Airport Security

Per the application of PRC, the authors initiated a study regarding airport security for the future SATS system.

Phase One – Policy Reviews

Congress has passed several important laws related to airport security over the past decade. From the electronic libraries of Congressional Universe and Lexis-Nexis, policy researchers are able to review current laws associated with the requirements, expenditures, and airport security development. Some essential laws related to airport security include:

1. Federal Aviation Act of 1958
2. Airport and Airway Development Act of 1970
3. Airport and Airway Improvement Act of 1982
4. Department of Transportation and Related Agencies Appropriations Act of 1990, 1996, 2001, 2002
5. Airport and Airway Safety, Capacity, Noise Improvement, and Intermodal Transportation Act of 1992
6. Federal Aviation Administration Act of 1994, 1996
7. Airport Security Improvement Act of 2000
8. Wendell H. Ford Aviation Investment and Reform Act for the 21st Century
9. Aviation and Transportation Security Act of 2001
10. Homeland Security Act of 2002
11. Vision 100 - Century of Aviation Reauthorization Act of 2003

These laws provide a legitimate base for airport security or relevant issues concerning personnel, facilities and equipment, trainings, budget, funding resources, and compliance deadline. Following the inspection of existing laws, policy researchers should review title 49 of United States Code (USC) for possible amendments of regulations in order to meet the regulatory needs.

Phase Two - Policy Research

In the discussion of SATS implementation, Step 4 would help identify the salient deficiency of aviation policy(ies) affecting current and possible future airmen or air operations. Step 5 (analytical findings) leads to possible policy mandates or new policy proposals (Step 6). For example, the Airport Security Improvement Act of 2000 amends the Section 44936(a)(1) of the United Stated Code (USC) Title 49 by adding "Criminal history record checks" at the end the subsection (E) for elevating the quality of screeners and compress the scope of overlooks based on personal discretion. The Act also enforces regular safety training and routine safety reports for all luggage screeners. Furthermore, the Aviation and Transportation Security Act of 2001 established the Transportation Security Administration (TSA), focusing on aviation security, safety, and related research and activities. By adding new duties and amending existing Title 49 of USC, the Aviation and Transportation Security Act empowers the agency/administrator to oversee airport security, federalize employees, train personnel, develop effective programs, and prevent potential threats. In the same vein, a review of Title 49 of USC would identify possible policy deficiencies related to SATS airport security such as the manpower qualification, security inspection training programs, funding resources, strategies for abnormal situations, the coalition among related agencies, and security report systems.

After identifying the possible policy deficiencies relevant to specific issues, researchers should look for answers to the following questions:
1) Is the policy or regulation inadequate, sufficient, or outdated respectively?
2) What would be the consequences if the policy remains unchanged?
3) What would be the possible impact on policy recipients?
4) How sound is the cost-benefit analysis related to the proposed policy?
5) Do airport users support such changes?

To answer these questions, policy researchers should not solely measure the possible outcome by forecast or statistics; in addition, expertise should be heard accordingly. Professional insight is very important when policy issues need special knowledge. Methods of collecting expertise include focus groups discussion, panel study, or Delphi techniques for those who cannot take part in discussions in person.

Phase Three – Policy Actions

Comparing the findings from statistical analysis and narrative expertise would draw researchers to answer a specific question: Do we need to propose a new policy or mandate existing laws related to airport security? If the answer is negative, policy researchers do not have to go through Stage 3, because current policies are able to sufficiently cover new challenges resulting from the operation of SATS. Conversely, if a policy mandate is necessary, policy researcher

should pilot-test the proposed mandate to a small voluntary group of SATS airports selected from industry. This procedure involves policy evaluation, revision, and implementation. During the testing period, policy researchers should also simultaneously collect feedbacks from the service providers/airports as well as the customers/passengers. This is simply because user contributions are the essential base for the government's decision making regarding the maximum interest of its people. The final stage of policy research will lead researchers to two choices: adopting or discarding of a suggested policy mandate. First, if the new mandate is adopted after pilot-tests and evaluation, the government would encounter less resistance from airports related to proposed policy change when it comes to nationwide implementation. Equally important, a policy implementation tracking system via benchmarking techniques is always plausible for measuring real-time needs associated with diverse changes. Second, if a new mandate is withdrawn, policy researchers would have to revisit the first stage of research construct and review the laws again. Or, researchers could return to Stage 2 for another in-depth dialog with related policy actors.

Conclusion

This proposed policy research construct not only tries to address the core of policy analysis and policy evaluation, it also tries to bridge the gap between two policymaking doctrines. It describes traditional decision-making pitfalls associated with empirical tools based primarily on statistics. "Randomly-selected samples" could not, to some extent, fully represent groups such as minorities or remote communities. Policymakers have utilized statistical laws and instrumental reasoning (i.e., regression analysis, correlation coefficient, Bayesian forecast, or t test) in policymaking for decades. However, academia has recently questioned this traditional (instrumental) policymaking reasoning. The authors also argued the accuracy of applied statistics for solving every legitimate problem. This debate has hedged policy development in aviation and provoked arguments from public administration (Fox & Miller, 1996; Rorty, 1982; Smith, 1998; White, 1999).

The introduction of PRC aims to broaden the decision-making range for aviation policymakers as well as to promote public participation in dealing with other policy issues. To date, a variety of aviation problems remain ineffectively addressed by policy researchers. In the wake of local communities' outcries about issues such as airport security, infrastructure, ground traffic, environmental impact, noise, and air pollution, the use of instrumental reasoning for aviation policymaking should be limited. As a result, SATS policy researchers must prepare to solve the aforementioned issues.

This article argues serious deficiencies exist in traditional policymaking by bureaucratic technocrats who prefer figures and linear procedures. Statistics-oriented or empirical policymaking is still popular, and the authors do not expect any research upheavals overnight in policy research methodology. Neither do

the authors seek to polemically attack or discard traditional policymaking procedures. Instead, the authors seek a more comprehensive policy research framework that can refine traditional analysis for lawmakers in hoping that they could provide public products more closely associated with the realistic needs of the air industry and the American flying public.

Final Comments

General readers may criticize the overall cost and time consumption regarding the discursive and reciprocal processes addressed in this research. However, to some extent, launching a cost-effect analysis seemed inappropriate because of the perceived value associated with a particular policy under this construct in terms of enhancing aviation safety and security. By *perceived value*, the authors mean that the utility (level of satisfaction) of decision-making could be grounded based on the particularity of a situation, which often presents a unique social value associated with the changing environment. Ontologically speaking, the high cost of enhancing aviation security and safety training was closely emphasized before the September 11th terrorist attacks. Yet concerning the costs of airport security and aviation safety is now a far less important issue. In this instance, the perceived value of a particular policy has seen its influence rise dramatically.

References

Bae, C. H. C. (1997). How Los Angeles' air policies benefit minorities. *Journal of Environment Planning & Management, 40* (2), 235-260.

Bernstein, R. J. (1983). *Beyond objectivism and relativism: Science, hermeneutics, and praxis*. Philadelphia, PA: University of Pennsylvania Press.

Blake, K. (2000). Reduced air travel delays? *Consumers' Research Magazine, 83* (4), 6.

Bowen, B. & Hansen, F. (2000). Implementing a national Small Aircraft Transportation System: A public construct. *Public Works Management & Policy, 5* (2), 160-169.

Bowen, B. & Tarry, S. (2001). Identification of public policy research application in the deployment of a new air transportation system component. *The Annual Meeting of the Transportation Research Forum* [CD-ROM]. Williamsburg, Virginia: Transport Research Board.

Bowersox, J. (2000). From water development to water management. *American Behavioral Scientist, 44* (4), 599-413.

Box, R. C. (1998). *Citizen governance: leading American communities into the 21th century*. Thousand Oaks, CA: Sage.

Centre for Research and Learning in Regional Australia (2000, June). *Policy-research forum—M forum outcomes*. University of Tasmania, Australia: Author.

Fox, C. J. & Miller, H. T. (1996). *Postmodern public administration: Toward discourse*. Thousand Oaks, CA: Sage.

Haas, P. J. & Springer, J. F. (1998). *Applied policy research: Concepts and cases*. New York, NY: Garland.

Hakim, C. (2000). *Research design: Successful designs for social and economic research*. New York. NY: Routledge.

Hutjes, J. (1991). Policy research between the accumulation and implementation of knowledge. *Knowledge & Policy, 41* (3), 10-26.

Jenkins, W. I. (1978). *Policy Analysis: A political and organizational perspective.* London: Martin Robertson.

Lu, C-t. (2003 May). An inductive study regarding a nonregulatory maintenance resource management (MRM) training. *Dissertation Abstracts International, 64(4),* 0704. (UMI No. 3087733).

Majchrzak, A. (1984). *Methods for policy research*. Beverly Hills, CA: Sage.

Niskanen, W.A. (1998). *Policy analysis ad public choice—Selected papers by William A. Niskanen.* Northampton, MA: Edward Elgar.

North, D. M. (1999). Aviation Week's action plan for air travel. *Aviation Week & Space Technology, 151* (17), 86-87.

Nyden, P., Figert, A., Shibley, M., & Burrows, D. (1997). *Building community: Social science in action*. Thousand Oaks, CA: Pine Forge.

Richardson, B. (2001, May 5). Old fuel and poor policy. *New York Times*, A13.

Rives, N. W.,Jr. & Serow, W. J. (1984). *Introduction to applied demography: Data sources and estimation techniques*. Beverly Hills, CA: Sage.

Robson, C. (1993). *Real world research: A resource for social scientists and practitioner-researchers*. Cambridge, MA: Blackwell.

Rogers, J. M. (1994). Organizational context, sponsorship and policy research output. *Knowledge & Policy, 7* (1), 3-24.

Rorty, R. (1982). *Pragmatism, relativism, and irrationalism*. United Kingdom: Cambridge University Press.

Schaaf, M. M. (2001). Air rage: a policy research study. *Collegiate Aviation Review, 19* (1), 168-184.

Scottish Council for Voluntary Organizations, Research Unite. (2001*). Policy research—*

A brief discussion. Retrieved September 30, 2001 from http://www.scvo.org.uk/research/

Smith, M. J. (1998). *Social science: In question*. Thousand Oaks, CA: Sage.

Tritter, J. (1995). The context of educational policy research: Changed constraints, new methodologies and ethical complexities. *British Journal of Sociology of Education, 16* (3), 419-430.

Walker, W. E. (1993). *Policy analysis and public infrastructure. Santa Monica, CA: Rand*.

Weiss, C. (1991). Policy research as advocacy: Pro and con. *Knowledge & Policy, 4* (1), 34-52.

White, J. D. (1999). *Taking language seriously: The narrative foundations of public administration research.* Washington, DC: Georgetown University Press.

Evaluation of a Spatial Disorientation Simulator Training for Jet Pilots

K. Wolfgang Kallus

and

Karin Tropper
Karl-Franzens-University
Dep. of Psychology
Universitätsplatz 2/III
A-8010 Graz, Austria
e-mail: wolfgang.kallus@uni-graz.at

Abstract

To evaluate the efficiency of a disorientation-recovery program using the Airfox DISO flight simulator (Disorientation Trainer, AMST Systemtechnik GmbH, Austria), 26 jet pilots were tested. The pilots were randomly allocated to 1 of 3 groups: Training, Awareness, or Control group. The training for the Training group included demonstration trials of the disorientation phenomena Gyro Spin, Leans, Dark Take-off, Expectation Error, Black Hole Approach, and False Horizon (awareness phase) and additional reorientation exercises (training phase). Pilots of the Awareness group also received the awareness phase, but instead of the training phase, they had the control condition free flight. All 26 pilots attended a test, which included profiles of disorientation elements. The whole procedure took about 7 hours per pilot. Aviation performance, psychological, and physiological data were measured. Results contribute to answer the more theoretical question how physiologically determined perceptual illusions and contradicting perceptual inputs to the mental picture of the situation interact to change or destroy situation awareness. Physiological stress occurred during the simulator exercises (e.g. Black Hole Approach) as indicated by significant changes in heart rate. Performance ratings show positive effects for the Training group. The Training group recovered more quickly, confirming these ratings.

Introduction

Spatial disorientation (SD) in flight is a frequently mentioned cause of accidents in military aviation (Bellenkes, Bason, & Yacavone, 1992; Braithwaite, Durnford, Crowley, Rosado, & Albano, 1998; Lyons, Ercoline, Freeman, & Gillingham, 1994). SD can be classified into three types: Type I (unrecognised spatial disorientation), Type II (recognised spatial disorientation), and Type III (overwhelming spatial disorientation) (Cheung, Money, Wright, & Bateman, 1995). Additionally, disorientation phenomena are classified into different groups according to the perceptual basis. Visual illusions are for example size-distance illusion, confusion of stars and ground lights, false horizon illusion, flicker vertigo or autokinesis. Somatogyral illusions are coriolis illusion, leans, or graveyard spin. Somatogravic illusions are the elevator illusion or the oculogravic illusion. There are also proprioceptive and geographical illusions. Spatial disorientation in jet pilots is due to perceptual illusions in most instances. A model to explain spatial disorientation in jet pilots is the mismatch-model, which states that input from different sources of information produces sensory conflicts (Bles, Bos, & Kruit, 2000). Due to a missing or insufficient explanation of the mismatch, recognised spatial disorientation will result. In other words, the mental model and the sensory input fail to match.

Considering systems with less speed like helicopters, the close links between spatial disorientation models and loss of situation awareness (Endsley, 1995; 1999) are evident. Situation awareness (SA) requires the correct perception, the correct interpretation of environmental stimuli, and the correct anticipation of future events. This is only possible as long as a match between the mental picture and the external situation is given. In dynamic systems, the future or projected mental picture plays a central role in SA. The future situation has to be predicted based on present situation cues, experience, and the mental model of the situation including the action and movement of the own system (Kallus, Dittmann, Van Damme, & Barbarino, 1999). Without a proper prediction and the match between prediction and the environment, our awareness quickly falls behind the dynamic of the situation. A theoretical model, which accounts for this anticipation-action-comparison-loop, is the model of Anticipatory Action Regulation (Hoffmann, 1993).

The goals of this study were the development and evaluation of a training program for disorientation phenomena in jet pilots. For this reason, reorientation training was developed based on simulations of the most frequent phenomena of perceptual illusions in jet pilots and these basic models of spatial disorientation. The study followed a psychophysiological multilevel assessment approach. Flight performance, physiological parameters, and mood state were complemented by behavioural ratings and interview data.

Method

Subjects and Design

Twenty-six male jet pilots (22 active jet pilots) in the age range of 24 to 60

years (mean = 33.5 years, s = 9.6) were randomly submitted to one out of three conditions. Every participant attended three flight simulator sessions (phases). The experimental design of the study is given in Table 1.

Table 1
Design

	Phase I	Phase II	Phase III
Training group (TG, n = 10)	Awareness	Training	Test1
Awareness group (AG, n = 8)	Awareness	Free Flight	Test1
Control group (CG, n = 8)	Free Flight	Test1 / Awareness	Test2

Ten subjects participated in the Training group, which consisted of an awareness phase and a training phase. The Awareness group, which consisted of eight pilots, had an identical awareness phase, but instead of the training, these pilots were allowed to practise in a free flight condition without intentional disorientation phenomena for the same amount of time. Pilots of the Training and the Awareness group attended test 1 during phase III. The eight pilots of the Control group had the free flight condition at the beginning (phase I). During phase II, they attended test 1 followed by the awareness exercises. During phase III, they had test 2, which was identical with test 1 (with one exception: the flight level for the recoveries in test 1 was 250 and in test 2 it was 100). The test 1 trials were identical for all groups and consisted of four disorientation profiles (Expectation Error, Dark Take-off, False Horizon, Black Hole Approach) and the profile Recoveries.

Flight Simulator DISO

The flight simulator Airfox DISO (Disorientation Trainer, Figure 1, Table 2) is a development of AMST Systemtechnik GmbH (5282-Ranshofen, Austria). An

external console is used to control the flight simulator. The instructor pilot (at the console) and the trainee-pilot (in the simulator unit) communicate via a radio connection. If necessary, the instructor pilot flies the simulator from the console, e.g. in order to set the flight parameters for the Recovery simulations (description cf. below). The profiles were developed for this study (Haug, 2002) and all active exercises used in this study were based on a F16 simulator.

Figure 1. Disorientation Trainer (DISO, http://www.amst.co.at).

Table 2

Kinematical parameters of the Airfox DISO (http://www.amst.co.at)

KINEMATICAL PARAMETERS			
Motion degree of freedom	Displacement	Velocity	Acceleration
Pitch	± 30 °	± 20 °/s	± 150 °/s2
Roll	± 30 °	± 20 °/s	± 150 °/s2
Yaw	± 60 °	± 20 °/s	± 150 °/s2
Additional yaw	360 ° continuous	150 °/s	± 15 °/s2
Heave	± 0,14 m	± 0,4 m/s	± 8 m/s2
Surge	+ 0,32 m / - 0,27m	± 0,4 m/s	
Sway	± 0,28 m	± 0,4 m/s	± 8 m/s2

Procedure

The awareness phase consisted of two passive Gyro Spin exercises and the five flight profiles Leans, Dark Take-off, Expectation Error, Black Hole Approach, and False Horizon. Four active exercises (1 x False Horizon, 2 x Black Hole Approach, 4 x Dark Take Off, and 3 x Expectation Error) were used for the training phase. In all profiles of the awareness phase and in some of the training phase some or all instruments (complete electrical failure) were switched off for a while for a better illustration of the disorientation phenomena. During the test phase (Expectation Error, Dark Take Off, False Horizon, Black Hole Approach and Recoveries) all instruments worked. For the profile Recoveries the instructor pilot set certain flight parameters via the external control console. During the set-up time the trainee-pilot (inside the simulator unit) kept his eyes closed. The trainee-pilot was instructed to reach safe flight parameters (to recover) as fast as possible. This exercise was conducted ten times according to the Recoveries listed in Table 3.

Table 3

Recoveries (same sequence for all pilots)

	Recoveries
1.	Nose High Recovery, VFR, high speed
2.	Inverted Recovery, VFR
3.	Nose Low Recovery, IFR, high speed
4.	Nose High Recovery, IFR, high speed
5.	Nose Low Recovery, VFR, high speed
6.	Inverted Recovery, IFR
7.	Nose High Recovery, VFR, low speed
8.	Nose Low Recovery, VFR, low speed
9.	Nose High Recovery, IFR, low speed
10.	Nose Low Recovery, IFR, low speed

The whole examination procedure lasted about seven hours per pilot. After a short briefing, the pilots were asked to fill in questionnaires. Before and after each phase in the DISO the pilots filled in an adjective check-list based on the German adjective check list of Janke & Debus (1978) and a physical symptom check-list based on the symptom list of Erdmann & Janke (1984) outside the flight simulator, and after each phase the pilots took part in an extensive reconstruction interview concerning the simulator profiles. The questions for analyzing mental processes were based on the Reconstruction Interview of the Integrated Task Analysis (ITA) that was developed to analyze mental processes in air traffic controllers (Kallus, Barbarino, & Van Damme, 1998).

Dependent Variables

Instructor Pilot Rating. The instructor pilot, an experienced jet pilot, who also maintained the radio communication with the participants, rated the flight performance of the pilots during all actively flown profiles along the evaluation criteria noted below:

- Allocation of Attention (AA)
- Situation Awareness (SA)
- Stress Resistance (SR)
- Multi Tasking (MT)
- Aggressiveness (AG)
- Overall Performance (OA)

As the Recovery sequences were of short duration (average about 12 sec.), the instructor pilot only rated the Overall Performance (OA). The four main ratings (evaluation categories) are excellent (4), good (3), fair (2), and unset (1).

Recovery Time. During the profile Recoveries (test 1 and 2) the Recovery Time was measured. The Recovery Time is the time it took the jet pilot to reach safe flight parameters (trainee-pilot's judgement) after the controls were handed over from the instructor pilot.

Psychical and physical state. The state of the subject was assessed by an adjective checklist and a physical symptom checklist before and after each phase in the flight simulator.

Heart Rate. Physiological stress reactions were assessed by ECG-registrations using a portable amplifier (g.tec, Graz) and a MATLAB/Simulink (Mathworks, Inc.) based software-system, which allowed online recording of the data via a standard notebook. The amplifier was installed behind the pilot's seat; the notebook used for data storage was also within the DISO unit. Furthermore, EEG and EOG were measured. To mark certain events (start of a profile, take-off, crash, etc.) triggers were set manually and they were recorded simultaneously with the physiological data.

Flight performance, psychological, and ECG data were submitted to statistical analyses using non-parametric (IP-Ratings) and parametric tests (Recovery-times and heart rate) from the SPSS statistical package. The analysis of the data follows the model of descriptive data analysis (Abt, 1987).

Results

Flight Performance

During the awareness phase, the instructor pilot ratings show no differences between the Training and the Awareness group. In contrast to this, descriptive analyses of the data of test 1 result in a better performance of the Training group compared with the Awareness or Control group on average. Significant training effects were obtained for the performance in the Expectation Error trial. (This flight is a night flight with stars and lights of a town visible. At the beginning of this profile, the F16 is already airborne at 3000 feet, 14 nautical miles from the airport. The pilot gets the instructions to come in for a Touch-and-Go manoeuvre and after this to fly a left-turn. During flying the left-turn, it is possible that the pilot becomes disorientated by not being able to discriminate between the lights of the town and the lights of the stars. Already during the approach, some pilots got similar orientation problems – as a result of the few lighted airport – as could be observed during the profile Black Hole Approach.)

Figure 2 shows the results for the categories Allocation of Attention (AA), Situation Awareness (SA), Stress Resistance (SR), Multi Tasking (MT), Aggressiveness (AG) and Overall Performance (OA); the Training group obtained higher scores (better ratings) in all categories.
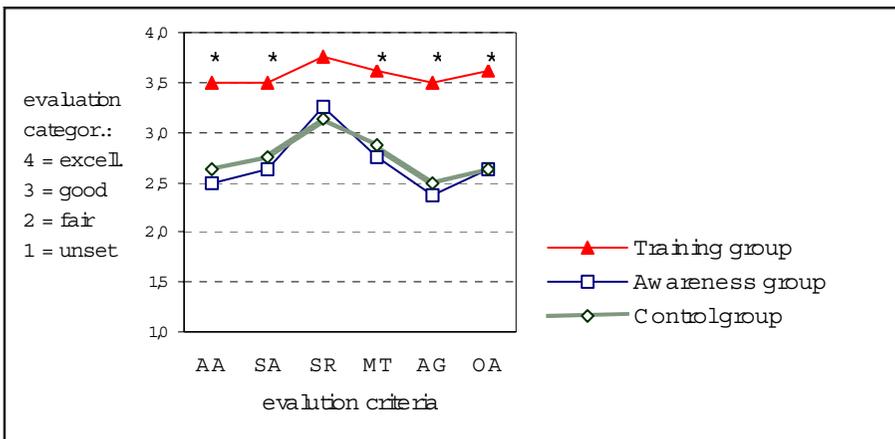
Figure 2. Performance evaluation by the instructor pilot (means) in the Expectation Error trial in test 1 per group (TG: n = 8, AG: n = 8, CG: n = 8), abbreviations cf. text.

Note that there are hardly any differences between the Control group and the Awareness group (Figure 2). Except stress resistance, all differences between

the Training group and the Control or Awareness group are significant ($p < 0.05$, Table 4).

Table 4
Results of the statistical analyses of flight performance ratings of the profile Expectation Error during test 1 (TG: n = 8, AG: n = 8, CG: n = 8)

	Kruskal-Wallis Test	Mann-Whitney U-Test	
Test1:	Exact p-value	Exact p-value	
Expectation Error	TG - AG - CG	TG - CG	TG - AG
Allocation of Attention (AA)	0,028	0,043	0,025
Situation Awareness (SA)	0,138	0,223	0,034
Stress Resistance (SR)	0,198	0,119	0,231
MultiTasking (MT)	0,019	0,010	0,010
Aggressiveness (AG)	0,020	0,034	0,015
Overall Performance (OA)	0,010	0,010	0,001

As can been seen in Figure 3 the Control group reached much better values in test 2 (phase III) than in test 1 (phase II). The differences are significant (Wilcoxon-test) for the evaluation criteria Situation Awareness (SA, $p = .038$), Stress Resistance (SR, $p = .025$), Aggressiveness (AG, $p = .038$) and the Overall Performance (OA, $p = .023$). Thus, attending test 1 with the subsequent awareness exercises during phase II resulted for the pilots of the Control group in a similar training effect to that obtained in the Training group.
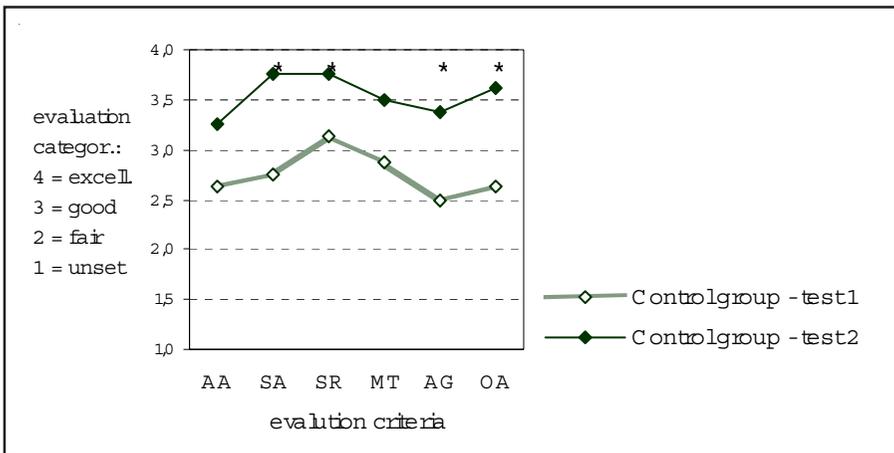
Figure 3. Changes in the performance of the Control group (n = 8) from test 1 to test 2, Expectation Error trial.

Ratings were supported by objective data. The Training group obtained best values i.e. lowest Recovery-times in nearly all recoveries (Figure 4). Significant

differences were obtained for Recovery trial 4 $(F(2,23) = 4.68, p = .020)$, for which pilots of the Training group had clearly lower recovery times than the Awareness group $(p = .016$, Post hoc tests: Tukey HSD).
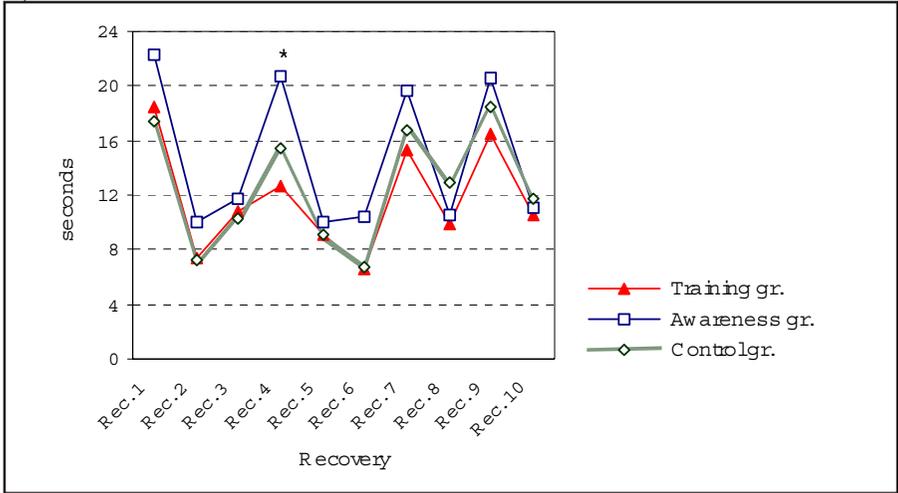
Figure 4. Recovery-times in test 1 for the ten different Recoveries per group.

While these results of the flight performance clearly show the effects of reorientation training procedures between the three groups, no significant group differences could be found for the self-rated psychological and physical states.

First analyses of the reconstruction interview data show that although the pilots were familiar with disorientation phenomena theoretically, they found themselves repeatedly in unexpected situations and their flight performance was often clearly below their personal expectations.

ECG
The physiological data show that disorientation profiles in a simulator can be quite stressful especially when disorientation occurs. As an example the Black Hole Approach, results of test 1 are given in Figure 5. (The Black Hole Approach is a night flight profile without peripheral visual cues, except a few-lighted airport. At the beginning of this profile, the F16 is already airborne at 3000 feet, 14 nautical miles from the airport. The pilot gets the instruction to come in for a landing. During the straight in approach disorientation phenomena can occur. As a result of the few-lighted airport the runway appears to move. Above this pilots get the visual illusion of a high-altitude final approach and if they believe their illusion – and don't look at or don't trust the instruments especially the altimeter – they decline too fast and too early which causes a crash in front of the runway).

For the analysis of the Black Hole Approach the pilots were – after flying this profile – assigned to three groups according to their flight performance: landing,

crash, and problems. "Problems" mean in this connection that the pilot had a bad landing (for example outside the runway), he did a Touch-and-Go manoeuvre or he decided to fly a Go-around.
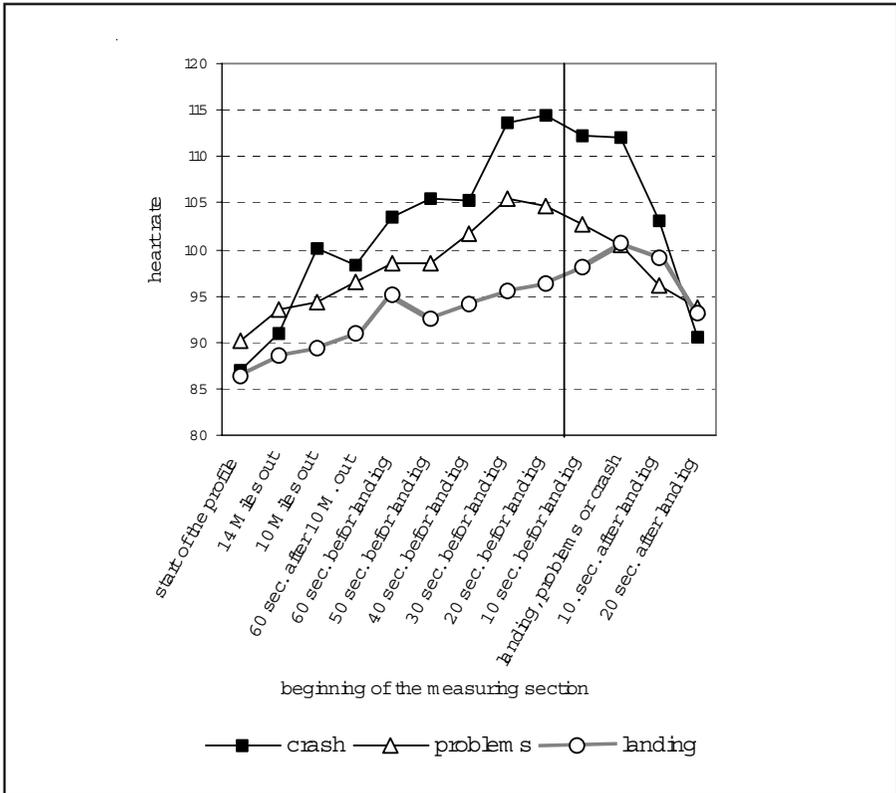
Figure 5. Changes in heart rate in the course of the profile Black Hole Approach for three different groups (crash: n = 6, problems: n = 6, landing: n = 10). To analyze the heart rate changes during this profile the complete physiological measurement from the start until 30 seconds after the pilot landed (or crashed, did a Touch-and-Go, etc.) was divided into 13 measuring sections. The labelling of the x-axis indicates the beginning of a measuring section.

Figure 5 shows that the standard landing (line with open bullets) shows the well-known and expected effect of a landing procedure. Disorientation (open triangles) and disorientation plus crash (filled squares) result in marked increases in heart rate (interaction time x group: $F(11.7,110.9) = 2,19$; $p = .018$; main effect time: $F(5.8,110.9) = 15.604$; $p = .000$).

<div align="center">Conclusions</div>

The results indicate that the Airfox DISO flight simulator illustrates disorientation phenomena realistically, although free movement (360°) is only possible along one axis and other movements are restricted to angles between 30° and

60° (cf. introduction, Table 1). The simulation scenarios are of high impact for the pilots, as could be demonstrated by the ECG-data. Technically, there are no problems in obtaining physiological recordings during the flight simulations. Thus, the Airfox DISO Simulator can be used to conduct detailed studies, which resolves the question whether the different disorientation phenomena follow a common mismatch-model. Research for motion sickness (Bles, 1998) indicated that there might be the need for a proper separation of different disorientation phenomena, as only certain events from the area of disorientation seem to cause motion sickness. On the other hand, simple perceptual illusions and complex mismatches in loss of situation awareness may have to be treated as different phenomena which nevertheless can be the cause of fatal "disorientation" errors in aviation.

It has been demonstrated that disorientation training for jet pilots should be done explicitly. Of course it is useful to improve awareness of disorientation phenomena in pilots with simulator profiles, as it is still a widespread phenomenon that the individual pilot believes that he personally will not be subjected to uncontrollable disorientation problems. Additional training – after the demonstration of disorientation profiles (awareness phase) – improves performance in dealing with disorientation remarkably. This may prevent fatal flight situations.

Recovery times were slightly increased for the Awareness group. This indicates that for anti-disorientation training a proper exercise is necessary to ensure a clear training effect. The need of proper application of new skills is well known from skill acquisition training and stress management training (Meichenbaum & Fitzpatrick, 1995).

The test profile Recoveries simulated situations in which pilots "recognise" the disorientation and try to recover. In contrast, the data shown for the Black Hole Approach include unrecognized flight problems. As heart rate increased for crash pilots in Figure 5 about 10 miles out, increased awareness about their state should have allowed a Touch-and-Go. This suggests that awareness of problem situations should be further increased by improved awareness training including "self awareness."

Current research addresses the generalization from jet pilots to VFR- and helicopter-pilots (with adapted profiles). If similar results are found, the current results are a strong argument for the inclusion of an anti-disorientation training in the training of pilots.

References

Abt, K. (1987). Descriptive data analysis: A concept between confirmatory und exploratory data analysis. *Methods of Information in Medicine*, *26*, 77-88.

AMST Systemtechnik. *Figure Disorientation Trainer and kinematic parameters*. Retrieved January 13, 2003, from http://www.amst.co.at.

Bellenkes, A., Bason, R., & Yacavone, D. O. (1992). Spatial Disorientation in Naval Aviation Mishaps: A review of class a incidents from 1980 through 1989. *Aviation, Space, and Environmental Medicine, 63*, 128-131.

Bles, W. (1998). Coriolis effects and motion sickness modelling. *Brain Research Bulletin, 47*, 543-549.

Bles, W., Bos, J. E., & Kruit, H. (2000). Motion sickness. *Current Opinion in Neurology, 13*, 19-25.

Braithwaite, M. G., Durnford, S. J., Crowley, J. S., Rosado, N. R., & Albano, J. P. (1998). Spatial Disorientation in U.S. Army Rotary-Wing Operations. *Aviation, Space, and Environmental Medicine, 69*, 10311037.

Cheung, B., Money, K., Wright, H., & Bateman, W. (1995). Spatial disorientation-implicated accidents in Canadian forces, 1982-92. *Aviation, Space, and Environmental Medicine, 66*, 579-585.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

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

Erdmann, G. & Janke, W. (1984). *Mehrdimensionale körperliche symptomliste (MKSL) [multidimensional scale on physical symptoms].* Berlin/Wuerzburg: University, Institute of Psyhology.

Haug, T. (2002). AMST SD Jet-pilot flight training. Airfox – aeromedical check-list*. Ranshofen: Unpublished Training manual.*

Hoffmann, J. (1993). *Vorhersage und Erkenntnis [Prediction and Recognition].* Göttingen: Hogrefe.

Janke, W. & Debus, G. (1978). *Die Eigenschaftswörterliste (EWL) [The Adjective Check List].* Göttingen: Hogrefe.

Kallus, K. W., Barbarino, M., & Van Damme, D. (1998). *Integrated task and job analysis of air traffic controllers phase 1: development of methods (HUM.ET1.STO1.1000-REP-03).* Bruessel: Eurocontrol.

Kallus, K. W., Dittmann, A., Van Damme, D., & Barbarino, M. (1999). Basic cognitive processes of air traffic controllers. In Harris, D. (Ed.) *Engineering Psychology and Cognitive Ergonomics Volume Three. Transportation Systems, Medical Ergonomics and Training* (pp. 113-120). Aldershot: Ashgate.

Lyons, T. J., Ercoline W. R., Freeman J. E. & Gillingham, K. K. (1994). Classification problems of U.S. Air Force spatial disorientation accidents, 1989-91. *Aviation, Space, and Environmental Medicine, 65*, 147-152.

Meichenbaum, D., & Fitzpatrick, D. A (1995). Constructivist narrative perspective of stress and coping: Stress inoculation applications. In L. Goldberger and A. Breznitz (Eds.), *Handbook of stress.* New York: Free Press, 1995.

Utilizing Preflight Observations to Facilitate the Development of Safety Cultures

Brian G. Dillman, Jang R. Lee, and Donald A. Petrin
Purdue University
West Lafayette, Indiana

Abstract
Many components of flight training that students subconsciously use to develop the safety culture, they will adopt. This mindset is acquired from the beginning of flight training and continues to grow throughout each student's respective career. There are many routine tasks that are accomplished for every flight that often become neglected over time, and this creates a gap in the safety culture that must be filled. One task that has the potential for becoming routine is the aircraft preflight. Aircraft preflight inspection is a learned process acquired early in the flight training of pilots. From these early lessons, it is assumed that a careful pilot will develop sound, enduring habit patterns for assessing the preflight airworthiness of various aircraft. However, because of differences in the instructional approach, variations in flight instructor experience, and individual student learning styles, observable, measurable differences in the way students approach the preflight process often result. Some students learn methodical, thorough, and consistent preflight methods while other students demonstrate slipshod, haphazard, and unsafe techniques. Clearly, assessment of aircraft airworthiness using the preflight inspection is the safety starting block for any flight. Because the instructional Law of Primacy suggests that what is learned first is best retained, it is imperative that proper, systematic preflight procedures be properly taught and judiciously followed in order to create and nurture both an individual acceptance of safety and a collective culture of safety.

Introduction

Before developing programs to create a safety culture, the concepts of "culture" and "safety culture" must first be understood. Since the actual measurement of culture is problematic, it is imperative that these guidelines be established. Kluckhohn and Strodbeck identified culture as a mega-variable from

which all contingencies originate. We think the way we have been socialized to think. The universal value orientations are to be found in all cultures. Each of the variations has implications for values, behavior, and emotions. Kluckhohn and Stodtbeck's five orientations are: (1) human nature, (2) relationships between man and nature, (3) time, (4) activity, and (5) relationships (Kluckhohn & Strodbeck, 1961).

One of the ways that an awareness of a safety culture can be promoted is by placing the idea of safety at the forefront, from the beginning of training. Analysis of selected aviation incident reports conducted by Orasanu (1997) suggested that situational awareness among crews was often compromised when workload was heavy or when distractions occurred. Consequently, it was concluded that strategies for managing crew tasks while promoting communication of essential information between crewmembers must be taught. Since preflight briefings are a crucial point for establishing the approach toward open communication, thus setting the tone for the remainder of the flight, this is precisely where the process for monitoring aircraft systems, enroute weather, and other related tasks should be formalized (Orasanu, 1997). A flight is a sequence of related events influenced by prior events. Since pilots, as a group, derive satisfaction from goal completion, once the commitment to flight is made, a compelling psychological incentive appears in the pilot psyche to continue to the intended destination (Smith, 1999). This sequence of events begins with the preflight inspection thus planting the seed for a psychological event that, in turn, will directly correlate to the quality of the safety culture.

Perhaps aviation safety should be approached as an ongoing, continuous process beginning early in primary training with the objective of creating a pilot society which recognizes the consequences of risk and which encourages individuals to form opinions based on a realistic knowledge of hazards (Thorburn, 1990). Unfortunately, most pilots do not receive structured decision or judgment either training in their initial or later flying experiences. It is assumed that they will learn judgment through their experience (Inagaki, Takae, & Moray, 1999). Thus, the final exam would come before any formalized awareness training. While it is true that a safety culture has no beginning and no end, conventional wisdom holds that the starting point for any given flight is the aircraft preflight. It is therefore imperative that sound processes be established, communicated, and reinforced to approach this early part of the flight with an appropriate sensitivity towards safety.

Statement of the Problem

According to Federal Aviation Regulation (FAR) 61.87 (d)(1), a student pilot must receive and log flight training for proper flight preparation procedures including preflight planning prior to solo flight. Since the FARs provide limited guidance as to the minimum training necessary for preflight, the quantity and extent needed to satisfy this requirement is left to the discretion of the individual

flight instructor. Perhaps paradoxically, these FAA minimums relate to the capabilities of ideal pilots and superb equipment because, in addition to regulating for safety, the FAA is also tasked with promoting the aviation industry (Jensen, 1997).

Within every discipline, there are teachers or instructors who are not just gifted presenters, but who also can connect with students in ways where students want to absorb everything possible. Their flight lessons, filled with rich information that students strive to absorb, go well beyond the FAA minimum requirements. For these superb, talented, and innovative individuals, aviation students with assorted aptitudes from various backgrounds will very likely succeed. Unfortunately, there also are those instructors who because of incompetence, lack of motivation, or negligence demonstrate extremely poor teaching skills and tolerate substandard learning. Sadly, because of the nature of aviation employment and the personal desire to amass flight time for job advancement, there are situations throughout the aviation training industry where flight instructors are teaching simply to increase flight time. Some instructors deliver only the barest minimum instruction required by the FAA so that they can get into the air with the hour meter running. These instructors may present topics in very broad terms, often glossing over the seemingly mundane skills such as preflight assessment of airworthiness. As a result, their charges are forced by circumstance to acquire their own paradigm of aviation safety through life experience. This teaching approach continues when students who were themselves improperly taught then become flight instructors, and subsequently imitate the way they were taught. A destructive cycle is thus embedded in the aviation industry, which measurably compromises the safety culture.

Since multiple flight instructors often teach in a flight program, it is necessary to establish rigorous standard operation procedures for many processes including the aircraft preflight. The importance of decisions made prior to flight was underscored by McElhatton and Drew (1993) in a study suggesting that many airline pilot decisions (or lack of decisions) made during the preflight phase directly correlated to incidents occurring much later in flight. Of the 125 Aviation Safety Reporting System (ASRS) incident reports reviewed in their study, ninety percent of all time-related human errors occurred during the preflight or taxi-out phase of operation (McElhatton & Drew, 1993).

Despite having formalized procedures in place for preflighting aircraft at Purdue University, there still are instances where because of time pressures, ignorance, sloth, or instructor inattention, students miss preflight items. These omission style behavior errors are usually trapped through a series of "safety nets" - a Swiss cheese approach where multiple assessments, observations, and safety consultations measure progress towards the individual safety component of the training process. Phase check evaluations at the completion of each flight course, certification flight tests for new certificates and ratings, daily observation, and a safety resolution process should change the behaviors of most unsafe students.

Yet, even though unlikely, it is still possible for someone to slip through these filters unnoticed and degrade the safety system.

The purpose of this paper is to determine whether standardizing the process of preflighting an airplane for flight students as a group would strengthen the process of establishing the culture of safety for a given flight.

Method

The hypothesis was that there were deficiencies in the preflight process of some of the students. Specifically, students did not know the proper procedure for preflighting an airplane, did not know which critical items that must be checked with regard to the engine compartment, and were doing too perfunctory of a preflight as a group. Furthermore, students were not carrying and referencing the preflight checklist, as was the standard operating procedure of the flight program. It was postulated that the amount of time spent on the preflight would increase following the discussion of preflight procedures between the control and experimental groups because more students would recognize the safety value of structured, consistent processes. As part of the experiment, wooden blocks were placed in the engine compartment blocking the inlet for the oil cooler (simulating a bird's nest) to see whether students were looking at specific areas of the engine compartment or simply completing a general sweep of the engine area. Care was taken to prevent students from starting the engine if they had overlooked the wooden obstacle.

The original structure for the observations of student aircraft preflights was to collect data on forty students as a control group, and forty students as an experimental group. Each of the subject students were enrolled in a collegiate flight program at a major Midwestern University. In order to randomize the selection of students for observation; the subjects were determined based upon the airplane tail number they chose rather than monitoring individual students. In other words, several aircraft would be parked on the flight line and the observer would determine which airplane provided the best view for collecting data. At this time the observer would wait until a flight student, of his or her own accord, would sign out a particular airplane and proceed to the flight line for preflight preparation. At this point data was collected on the preflight process. While students might have a bias for a particular airplane, parking spot, or flight period, the process for subject selection eliminated the potential bias of the observer for selecting those students that would give the best data. It was the elimination of the observer bias, for which the subject selection process was chosen.

Subject information such as student name, flight period, and aircraft number were recorded in order to prevent multiple observances of the same student within and between the control and experimental groups. Prior to analyzing the data, subject information was compared and it was determined whether or not there were duplicate observations. The determination was made that there were no duplicate observations within or between the control and experimental group.

Information such as total time of preflight, time of interior and exterior pre-flight, time of preflight spent in the engine compartment, and the use of preflight checklists was to be collected without subject knowledge. A series of follow-on questions (Appendix A) were asked to determine the student's perceived thor-oughness of the process used to preflight the aircraft. After an individual was observed as part of the control group, they were asked to not share information concerning the study with any other students to preclude the possibility of sample contamination between observed and non-observed students. After a few more than twenty observations for the control group it was determined that prior stu-dents had indeed shared information about the study and the decision was made to stop the collection of data for the control group at twenty-two. This decision was made after the final observed student in the control was asked, "What do you look for in the engine compartment?" The response was, "Birds nests and wooden blocks."

After the data for the control group was collected, all of the flight students were brought together and standardized as to the desired process used to pre-flight a Piper Warrior III. Procedures established by the target university, includ-ing examining maintenance records, utilization of interior/exterior checklists (items critical to flight safety that must be checked during each preflight) were reviewed with each flight student.

After each flight student was presented with the accepted procedure for preflighting the Piper Warrior III, forty students were observed for the experimen-tal group and information was now collected *after* the flight to determine the thoroughness of the preflight process. The same information concerning length of time for preflight and a series of questions (Appendix B) were asked of each student in the experimental group.

Observation Results

The original hypothesis was that students were not completing thorough preflights and therefore starting the process of safety for a given flight with an inappropriate safety paradigm. Consistently starting the process of safety for a given flight in an incorrect manner leads to a breakdown of the overall safety culture for flight students. The expected result of the project was an increase in the amount of time being dedicated to preflights and consistency in the proce-dures being utilized for the preflight process following standardization of the preflight process. The following data for the control and experimental groups can be found in Appendix C and D with figures comparing the two groups in Appen-dix E.

The control group had an average overall preflight time of 10:01 minutes and seconds while the experimental group had an average of 8:35 minutes and sec-onds. This was a *decrease* in the amount of overall preflight time of 1:26 minutes and seconds or about 14.3%.

The control group had an average time spent in the cockpit of 2:52 minutes and seconds while the experimental group had an average of 3:00 minutes. This was an *increase* in the amount of time spent in the cockpit of :08 seconds - an increase of 4%.

The control group had an average time spent in the engine compartment of 1:43 minutes and seconds while the experimental group had an average of 1:42 minutes and seconds. This represented a *decrease* in the amount of time spent in the engine compartment of :01 seconds, which is less than 1%.

The control group had an average time for the exterior walk around of 6:16 minutes and seconds while the experimental group had an average of 5:26 minutes and seconds. This was a *decrease* in the amount of time spent during the exterior walk around of 13.3%.

With regard to following proper preflight procedures, the control group contained 12 students who carried a physical checklist while 10 did not carry the checklist. Nine of the control group students actually referenced the checklist while 13 did not. Thus, for the control group, 54.5% of the students carried the checklist and 40.9% of the students referenced the checklist during the exterior walk around. The experimental group produced 37 students who carried the physical checklist and 3 who did not. Thirty-one of these students referenced the checklist while nine did not. Accordingly, for the experimental group, 92.5% of the students carried the checklist while 77.5% of the students referenced the checklist during the exterior walk around. This represented an *increase* of 38% for carrying the checklist and 36.6% for referencing the checklist from the control to the experimental groups.

Thoroughness of the student preflight inspections was assessed using perceived fuel on board and date of the next required inspection. Using 'the amount of time before the next required maintenance inspection", the control group had 5 students who did not know, 4 who had a general idea, and 12 who knew specifically, which is a success rate of 72.7%. By comparison, the experimental group had 37 students who knew specifically and 3 who did not fly for various unrelated reasons but were unable to state the amount of time prior to the next required inspection - a success rate of 92.5%. This represents an *increase* of 19.8%. Using "the amount of fuel on board converted into flight time available," the control group had 9 who knew the exact time in hours and minutes and 13 who only knew how many gallons of fuel were available - a success rate of 40.9%. The experimental group had 36 who knew the exact amount of time and 3 who did not fly for unrelated reasons but were unable to state the amount of "time" in the fuel tanks, which is a success rate of 92.5%. This represents an *increase* of 51.6%.

Discussion

There are several conclusions that may be drawn from the data collected

from the control and experimental groups. The original hypothesis was that an increase in the amount of time for the preflight would result following the standardization of the preflight process. The reasoning behind this hypothesis was that a more thorough preflight would take more time and it was expected that the control group was not performing a sufficiently thorough preflight. After examining the data, it became apparent that this was not the case. The average time spent during the various segments of the preflight for the experimental group *decreased* for all except the interior preflight segment. After several student pilots had completed the preflight, this trend was indeed observed. Although the information was not consistently obtained from each of the preflight subjects, an informal survey was completed by each participant to quantify the reason for this decrease in the amount of time dedicated to the preflight process.

Answers obtained from the informal survey suggested that students had become more efficient during the preflight due to the standardization process. Several students commented that they, "Now knew what to look for during the preflight." Before the standardization process was completed, students were randomly examining various aircraft components with little justification for the order of their actions. After the standardization process was completed, the students were equipped with specific knowledge on how to complete a thorough and accurate aircraft preflight.

A second observation concerned the procedures that are followed during the preflight process. Standard operating procedures required students to carry and reference a preflight checklist during the exterior walk around. This standardized approach was crafted so that various segments of the aircraft were not overlooked even if students were interrupted by ramp events, conversation with other students, or flight instructor questions.

The percentage of students who carried the exterior preflight checklist *increased* from 54.5% to 92.5% after the standardization process was completed. This represented a significant increase in the number of students following the correct procedures.

All flight departments have standardized procedures that are expected to be followed by all pilots on the flight line. The percentage of those pilots following the procedures can be greatly increased by two simple factors. First, provide each pilot with the standardized procedure and explain in detail what behavioral outcomes are expected during the process. Second, and perhaps more importantly, there must be a system of checks to ensure that the standardized process is being judiciously followed. One aspect of flight training that this project has emphasized is the need for random checks of standard operating procedures in the Operations Manual.

During initial flight training, students are taught to complete a preflight by a certified flight instructor and then students are observed for several consecutive preflights until a pattern of consistent proficiency is seen. Following this early demonstration of proficiency, it might be assumed that the student is indeed competent in the preflight process and that further observations are unnecessary. Yet, with the passage of time, student preflight may erode into a superficial process where thoroughness learned a priori will suffer. By contrast, using randomized observations, the students will soon recognize that they are constantly under evaluation and as a result, the quality of their preflight assessments will remain high. This "striving" behavior is exactly what was observed during data collection. The number of students in the control group represented those students who had accepted a certain level of indifference in the preflight process. After the standardization process was completed, and the realization that preflights were now being observed; the students re-adopted the established, standardized procedures and also carried the required physical checklist as specified in the Flight Operations Manual.

Moreover, the percentage of students referencing the checklist *increased* from 40.9% to 77.5%. Although this final percentage is not as high as the number of students carrying the checklist, it still represents a significant increase from the control group. Once again, as students complete multiple preflights on identical aircraft, a level of complacency may develop that must be overcome. If a pilot completes 100 preflights over the course of several months and the result is always the same, there is a human tendency to expect the same level of airworthiness on subsequent observations. Using randomized observations to collect the project data forced students back to appropriate standardized procedures, which, in turn, improved the level of individual safety and enhanced the collective safety culture.

The final basis for quantifying the thoroughness of the preflight process was to evaluate the student on determining the amount of time before the next required inspection and the amount of fuel on board converted into hours of flight time. These two criteria were chosen because of past experiences where students flew airplanes past inspection times and landed airplanes with significant imbalances between fuel tanks. The percentage increase from the control to experimental groups for determining time prior to inspection and fuel on board was 19.8% and 51.6% respectively. Even though the percentage increase of students who knew the amount of "time prior to the next required inspection" was not as significant as the percentage for "fuel on board," the answer accuracy to the "time prior to next inspection" tells a different story.

The experimental group gave more accurate answers than did the control group, and, in many cases, knew exactly how much flight time until the next required inspection remained to the nearest 10th of an hour. Several of the control group students had only a general idea as to how much time was left and in some cases did not know at all. Of the experimental group, the only students

who were not aware of the exact amount of time prior to the next inspection were those students who did not fly the airplane for various reasons. Clearly, the accuracy of the answers obtained from the experimental group concerning the amount of fuel on board also was significantly improved over the control group. The experimental group knew precisely how many gallons of fuel were on board and the amount of flight time that fuel equated to for a particular operation. Several of the control group subjects only determined how much fuel was on board the aircraft and not how much flight time it would provide. Although this may seem like a minor difference, in the aviation world, it can be the difference between properly balanced fuel, landing with adequate and safe reserves or running one or more tanks totally dry. Each year there are several accidents where improper fuel management is the root cause. Because pilots cannot equate miles per gallon with consistent accuracy as in the case of automobiles, they must utilize a different measurement to determine fuel on board. If a pilot knows that there are four hours of fuel on takeoff, then at any given point in time, they would have an accurate estimate of how much longer the airplane will fly before becoming a glider.

Conclusions

It is the seemingly insignificant steps such as converting fuel on board into flight time, consistent thoroughness of the preflight inspection, and a professional approach to each flight that supports a culture of safety. Periodic observations to determine whether standard operating procedures, such as aircraft preflight, are valuable to keep flight students and instructors on the right track. Clearly, the "ownership" of aviation safety is a notion that must be implanted within the psyche of each individual pilot from the very first lesson during the preflight process.

References

Federal Aviation Administration. (2003). *14 C.F.R § 61.87*. Retrieved December 15, 2003, from http://www.access.gpo.gov/nara/cfr/cfrhtml_00/Title_14/ 14cfr23_00.html

Inagaki, T., Takae, Y., & Moray, N. (1999). Automation and human-interface for takeoff safety. In R.S. Jensen, B. Cox, J.D. Callister, & R. Lavis (Eds.), *Proceedings of the Tenth International Symposium on Aviation Psychology*. (pp. 402-407). Columbus, OH: Ohio State University Press.

Jensen, R. (1997). Risk management in general aviation: Making your own rules. In R.S. Jensen & L.A. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology*. (pp.727-732). Columbus, OH: Ohio State University Press.

Kluckhohn, F. R., & Strodbeck, F.L. (1961). *Variations in value orientations*. Westport, CT: Greenwood Press.

McElhatton, J. & Drew, C. (1993). Time pressure as a causal factor in aviation safety incidents – The "hurry-up" syndrome. In R. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 269-274). Columbus, OH: Ohio State University Press.

Orasanu, J. (1997). Training situation awareness in aviation. In R.S. Jensen & L.A. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology*. (pp.1462). Columbus, OH: Ohio State University Press.

Smith, K. (1999). Incompatible goals, uncertain information, and merciless incentives: The dispatch dilemma. In R.S. Jensen, B. Cox, J.D. Callister, & R. Lavis (Eds.), *Proceedings of the Tenth International Symposium on Aviation Psychology*. (pp.445-451). Columbus, OH: Ohio State University Press.

Thorburn, S. (1990). Safety through education and training. *European Journal of Engineering Education*, 15(1), 13-19.

```
APPENDIX A
Control Group

                  Safety Culture Observation – Control Group
                            Aircraft Preflight


Observation Data:


Amount of time spent in cockpit
Amount of time spent in engine compartment
Amount of time for walk around
Amount of time spent on entire preflight
Was a physical checklist carried during exterior preflight?
Was a physical checklist referenced during exterior preflight?
Did the pilot start in one position and move around the airplane in a logical
sequence?



Questions asked of pilot:


How much time before next inspection?
How much flight time do you have in the fuel tanks?
Did the pilot find the "wrench left by maintenance" in the engine compartment?


            If not found:


            What do you look for in the engine compartment?
```

```
APPENDIX B
Experimental Group

               Safety Culture Observation – Experimental Group
                            Aircraft Preflight


     Tail#                        Time                        Date

Observation Data:


Amount of time spent in cockpit
Amount of time spent in engine compartment
Amount of time for walk around
Amount of time spent on entire preflight
Was a physical checklist carried during exterior preflight?
Was a physical checklist referenced during exterior preflight?

Questions asked of pilot:

How much time before next inspection?
How much flight time do you have in the fuel tanks upon landing?
```

APPENDIX C
Control Group Data

Ob #	Time spent in cockpit	Time in engine	Time for walk around	Total time for preflight	Physical checklist carried	Physical checklist referenced	Preflight in logical sequence	Time before next inspection	Time in fuel tanks	Find block?	Look for in engine
1	1:30	0:40	2:30	4:40	NO	NO	YES	DON'T KNOW	FULL	NO	OBSTRUCTION
2	2:00	0:42	4:12	6:54	YES	NO	YES	DON'T KNOW	5 HRS	YES	
3	3:00	0:45	7:00	10:45	NO	NO	YES	100	18 GAL EACH	NO	
4	3:20	1:30	9:00	13:50	YES	YES	YES	NOW	3.5	YES	
5	2:10	2:20	7:00	11:30	YES	YES	YES	10	FULL	YES	
6	1:10	0:35	3:15	5:00	YES	NO	YES	40	FULL		
7	3:20	0:40	6:20	10:20	YES	YES	YES	30	FULL		
8	2:00	5:00	12:10	19:10	YES	YES	YES	DON'T KNOW	FULL		
9	2:20	2:00	7:40	12:00	YES	YES	YES	4	5		
10	1:15	1:15	5:00	7:30	NO	NO	YES	13	FULL	YES	
11	2:25	1:15	3:56	7:36	NO	NO	YES	50	4		
12	3:00	1:50	5:10	10:00	NO	NO	YES	40	5		
13	3:15	2:47	9:30	12:45	YES	YES	YES	GENERAL OK	FULL	NO	LOOSE CONNECTIONS, LEAKS, FLUIDS
14	4:04	1:47	8:17	12:21	YES	YES	YES	PLENTY	FULL	NO	CYLINDER, LEAKS, CONNECTIONS
15	4:02	1:50	7:34	11:36	NO	NO	YES	15	4	NO	FLUIDS, DRIPS, SMELLS
16	3:55	1:39	6:45	10:40	YES	NO	YES	JUST DONE	5	NO	CONNECTIONS, LEAKS
17	3:12	0:49	2:08	5:20	NO	NO	YES	NOT YET CHECKED	FULL	NO	LEAKS, CONNECTIONS
18	3:15	1:17	6:10	9:25	YES	YES	YES	NOW	4	NO	LEAKS, FLUIDS
19	5:20	1:30	4:55	10:15	NO	NO	YES	NOW	4	NO	GENERAL PROBLEMS, THINGS THAT SHOULDN'T BE THERE
20	2:20	3:10	7:35	9:55	YES	YES	YES	GENERAL OK	FULL	NO	OIL, LEAKS, FLUIDS, CONNECTIONS
21	3:18	2:51	5:29	8:47	YES	YES	YES	7 OR 8	FULL	YES	OIL, FLUIDS, P-LEAD RAGS, ALT BELT
22	3:15	3:00	6:05	12:20	NO	NO	YES	39	FULL	NO	BIRDS NESTS AND WOODEN BLOCKS
AVE	2:52	1:43	6:16	10:01	12YES	9YES	22YES	5 DON'T KNOW	9 KNEW EXACT TIME	5 FOUND BLOCK	
					10NO	13NO		4 GENERAL IDEA	13 KNEW HOW MUCH FUEL	11 MISSED BLOCK	
									12 KNEW SPECIFI-CALLY	7 DIDN'T HAVE BLOCK	

APPENDIX D
Experimental Group Data

Ob #	Time spent in cockpit	Time in engine	Time for walk around	Total time for preflight	Physical checklist carried	Physical checklist referenced	Time before next inspection	Time in fuel tanks
1	4:35	1:40	4:05	8:40	YES	YES	41	3.5
2	6:30	2:13	6:55	13:25	YES	YES	Did not fly	Did not fly
3	4:20	2:38	7:05	11:25	YES	YES	21.6	4
4	5:05	1:47	7:05	12:10	YES	YES	13	2.5
5	4:50	2:05	8:40	13:30	YES	YES	0	3.5
6	3:40	1:35	8:20	12:00	YES	YES	38	2.5
7	4:18	1:30	4:55	9:13	YES	NO	18	3:10
8	2:00	1:05	4:20	6:20	YES	YES	26	3.5
9	2:57	2:30	5:08	8:05	YES	YES	31.5	3
10	3:28	0:35	2:25	5:53	YES	NO	18.1	3.85
11	3:50	1:25	4:50	8:40	YES	YES	14.8	2:48
12	3:30	0:55	4:20	7:50	YES	NO	28.5	3.5
13	3:25	1:20	5:02	8:27	YES	YES	27.1	4
14	3:15	3:05	6:05	9:20	YES	YES	14.5	3.5
15	2:20	0:50	8:45	11:05	YES	YES	0	3.4
16	2:30	1:55	7:50	10:20	YES	YES	23.4	3.5
17	2:25	0:55	4:45	7:10	YES	NO	43	2.9

APPENDIX D - continued
Experimental Group Data

18	1:10	0:45	2:35	3:45	YES	NO	41	3.9
19	3:15	1:20	5:25	8:40	YES	YES	43	4:10
20	3:20	1:30	4:00	7:20	YES	YES	Did not fly	Did not fly
21	3:19	3:13	9:24	12:43	YES	YES	3.7	1
22	3:19	3:03	7:57	11:16	YES	YES	32	26gal
23	5:37	2:17	8:22	13:59	YES	YES	19.2	4
24	2:02	2:57	5:52	7:54	YES	YES	39	3.1
25	2:47	1:55	5:02	7:49	YES	YES	7	4
26	1:45	1:59	5:00	6:45	YES	YES	27	4.1
27	2:24	1:50	3:28	5:52	NO	NO	34	6
28	1:54	1:33	4:02	5:56	YES	YES	35	3.5
29	2:07	0:48	2:21	4:28	YES	NO	0	4.5
30	2:06	1:29	4:39	6:45	YES	YES	39	3.4
31	1:18	1:21	9:21	10:39	YES	YES	32	4
32	1:55	1:27	4:55	6:50	NO	NO	23	2.5
33	2:42	1:38	4:37	7:19	NO	NO	19	3.5
34	1:45	0:58	4:12	5:57	YES	YES	29.7	3.1
35	2:15	1:33	5:00	7:15	YES	YES	34	3.6
36	2:28	1:23	3:52	6:20	YES	YES	48	3
37	1:56	1:51	5:42	7:38	YES	YES	30	2.5
38	2:28	2:01	5:35	8:03	YES	YES	Did not fly	Did not fly

APPENDIX D - continued
Experimental Group Data

								37 Knew exact time
AVE	3:00	1:42	5:26	8:35	37 YES	31 YES	37 Knew Specifically	

APPENDIX E
Comparison of Control and Experimental Groups

Time Spent in Cockpit — 4% Increase

Time for Exterior Walk Around — 13.3% Decrease

Time in Engine Compartment — < 1 % Decrease

Total Time for Preflight — 14.3% Decrease

Physical Checklist Carried — 38% Increase

Physical Checklist Referenced — 36.6% Increase
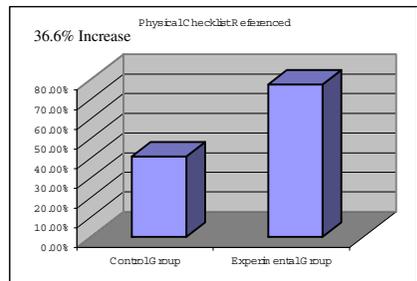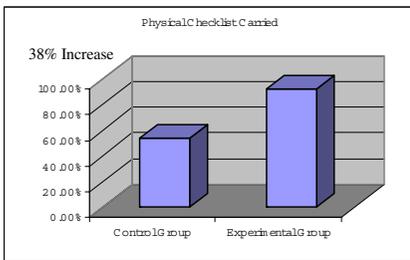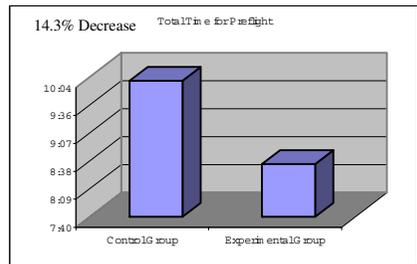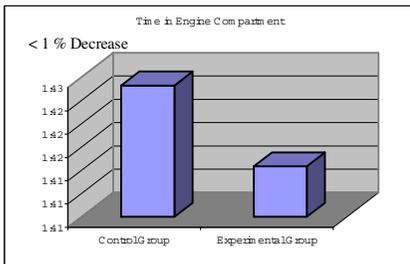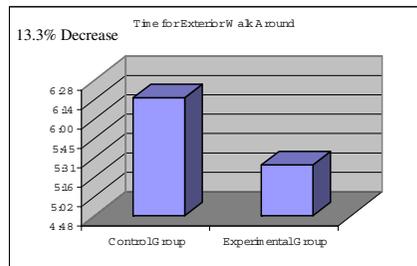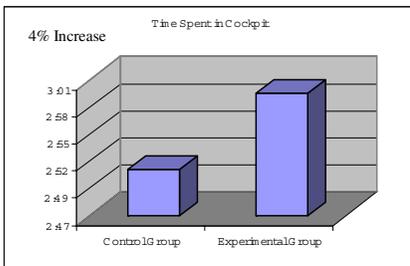
APPENDIX E - continued
Comparison of Control and Experimental Groups
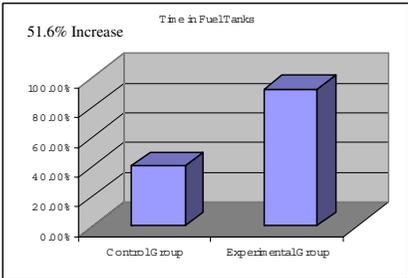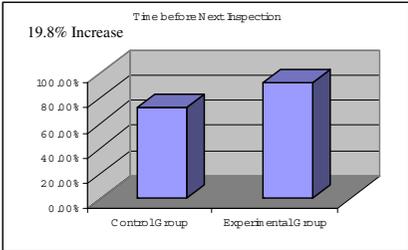
The Role of Organizational and Individual Variables in Aircraft Maintenance Performance

Gerard J Fogarty
University of Southern Queensland
Toowoomba, Queensland 4350
Australia
Email: fogarty@usq.edu.au

Abstract

Aviation maintenance has been identified by the FAA as an area where better efficiency is needed to cope with ever increasing workloads. However, aviation maintenance also has been identified as one of the major causes of accidents. Consequently, if further efficiencies are to be achieved, they cannot come at the cost of reduced safety margins. The present study employed a safety climate approach to assist in the development of a model that can help to explain morale, psychological health, turnover intentions, and error in the aviation maintenance environment. An instrument called the Maintenance Environment Survey was developed and administered to 240 personnel responsible for maintenance of a large military helicopter fleet. Data collected through the survey were used to develop a structural model that predicted 45% of the variance in psychological health, 67% of the variance in morale, 27% of the variance in turnover intentions, and 44% of the variance in self-reported maintenance errors. The model shows the pathways through which organizational level and individual level variables can influence work outcomes and leads to suggestions for interventions that can help to improve maintenance efficiency.

The Role of Organizational and Individual Variables in
Aircraft Maintenance Performance

The importance of the maintenance function was captured by Weick and colleagues when they observed that: "Maintenance people come into contact with the largest number of failures, at earlier stages of development, and have an ongoing sense of the vulnerabilities in the technology, sloppiness in the operations, gaps in the procedures, and sequences by which one error triggers another" (Weick, Sutcliffe, & Obstfeld, 1999, p. 93). A significant proportion of these errors come at the hands of the maintainers themselves as the ever-increasing complexity of aviation places greater demands on those responsible for their maintenance.

Figures emerging from the United Kingdom Civil Aviation Authority (CAA) showed a steady rise in the number of maintenance error mandatory occurrence reports over the period 1990 to 2000 (Courteney, 2001). A recent Boeing study of worldwide commercial jet aircraft accidents over that same period showed a significant increase in the rate of accidents where maintenance and inspection were primary factors (cited in ICAO, 2003). The FAA, in its strategic plan for human factors in aviation maintenance through to 2003, cited statistics from the Air Transport Association of America (ATA) showing that the number of passenger miles flown by the largest US airlines increased 187% from 1983 through to 1995. Over that same period, the number of aircraft operated by those airlines increased 70%, but the number of aviation maintenance technicians increased only 27%. The FAA concluded that the only way the maintenance program could cope with the increased workload was by increased efficiency at the worker level (cited in McKenna, 2002).

Despite the awareness of the importance of maintenance to the aviation industry and the growing problems confronting maintenance, until recently, empirical research into the nature of maintenance work and related human factors has been negligible. The development of descriptive models of human error and accident causation (Reason, 1990; Senders & Moray, 1991) and the recent adaptation of Reason's model to aviation maintenance (Reason & Hobbs, 2003) are major steps in the right direction. Research on error classification schemes (e.g., Patankar, 2002; Shappell & Weigmann, 1997) and, more recently, safety culture (Taylor & Thomas, 2003; Patankar, 2003) represent other bright spots in a surprisingly sparse research literature. However, what are needed, in addition to the descriptive accident causation models, classification schemes, and culture surveys, are empirically validated models that capture the major influences on maintenance work and provide a means of assessing these influences. Models of this kind can provide the basis for predicting unsafe organizational states and designing interventions that will lead to reductions in maintenance errors. The present study set out to develop such a model within the context of aviation maintenance using a multivariate methodology that has its roots in what has become known as the safety climate approach. This approach is described in the following paragraphs.

Safety Climate and Safety Culture

Over the years, the concepts of safety culture and safety climate have developed almost in parallel through the safety literature. Safety climate is operationalised in the current study as the individual's perceptions of the organizational policies, procedures, and rewards relevant to safety in the organization (Guldenmund, 2000; Griffin & Neal, 2000). This definition sets it apart from safety culture, which is usually regarded as a stable, deep-seated aspect of an organization that expresses itself through climate (Guldenmund, 2000, p. 221). Whereas the assessment of safety culture requires tangible means of measurement such as in-depth interviews and analysis of stated safety goals and polices (Guldenmund, 2000; Mearns & Flin, 1999), safety climate is assessed through self-report questionnaires.

Constructing a Measure of Safety Climate

Attempts have been made to define a core set of constructs for safety climate (see Flin, Mearns, O'Connor, & Bryden, 2000). Although not entirely successful in establishing core dimensions, this research is useful in suggesting constructs that should be considered for inclusion in research on maintenance errors. Recent publications relating to the assessment of safety climate in aviation maintenance also provide guidance. Taylor and Thomas (2003), for example, used a self-report questionnaire called the Maintenance Resource Management/ Technical Operations Questionnaire (MRM/TOQ) to measure what they regarded as two fundamental parameters in aviation maintenance: professionalism and trust. The dimension of professionalism is defined in their questionnaire in terms of reactions to work stressors and personal assertiveness. Trust is defined in terms of relations with co-workers and supervisors. Questions relating to these areas also appear in the questionnaire to be used in the current research. Patankar (2003) constructed a questionnaire called the Organizational Safety Culture Questionnaire, which included questions from the MRM/TOQ along with items from questionnaires developed outside the maintenance environment. Following the application of exploratory factor analytic routines to a dataset generated from respondents that included 124 maintenance engineers, Patankar identified four factors as having particular relevance to the safety goals of aviation organizations: emphasis on compliance with standard operating procedures, collective commitment to safety, individual sense of responsibility toward safety, and a high level of employee-management trust.

Turning to the general safety literature, there are now a host of questionnaires that purport to measure either safety culture or safety climate. Wiegmann and his colleagues (Wiegmann, von Thaden, Mitchell, Sharma, & Zhang, 2003) drew upon 13 such measures to construct their Commercial Aviation Safety Survey (CASS), an instrument designed for use with pilots. Most of these questionnaires are multidimensional, covering a range of factors that the authors consider to be of relevance to safety performance. The availability of so many questionnaires tapping an array of safety-related constructs presents a challenge to researchers interested in constructing a safety climate survey for use in specific settings such as maintenance.

That challenge was addressed in the present study by using the principle of triangulation to isolate the constructs relevant to a maintenance environment. Drawing upon the distinction between culture and climate made earlier, this methodology entailed a close examination of the safety culture in an organization in order to derive questions for inclusion in a safety climate survey. The first step in the triangulation process involved a search of the safety literature to identify potential constructs for inclusion in the questionnaire. As already mentioned, there is no shortage of surveys in the literature and some researchers have attempted to identify core safety climate constructs (e.g., Flin et al., 2000). The second step involved the analysis of a maintenance incident database and the associated incident investigation reports. The database and incident reports highlighted the relevance of factors such as inadequate training, poor supervision, and individual factors such as stress and fatigue as causes of maintenance-related incidents. The third method involved a series of focus group interviews with maintenance personnel and their supervisors to ascertain their perceptions of factors that impact on maintenance work. Content analyses of these interviews highlighted organizational concerns such as scheduling and resources.

Information collected in these three phases was then used as the basis for the construction of a questionnaire to measure organizational and individual factors considered likely to impact on maintenance performance. The resulting questionnaire, called the Maintenance Environment Survey (MES), was broader in scope than many of the existing climate or culture surveys. It contained questions intended to define the following constructs: a) safety climate, b) morale, c) psychological health, d) job turnover intentions, and e) maintenance errors.

The construction and validation of the MES was a necessary first step towards the development and validation of a structural model showing how the various factors captured by the survey interact to influence maintenance errors. Despite the proliferation of studies reporting new safety climate questionnaires, there are few studies in the safety literature that have taken the extra step of constructing models to illustrate the interactions among the psychological factors captured by the questionnaires. Using climate surveys in combination with the techniques of multivariate analysis, especially path analysis and structural equation modeling, it is possible to capture elements of the accident causation process and to test different models of how the components of the system work. These models can then be used to direct interventions aimed at improving safety performance in the maintenance environment. The rationale for the model to be tested in the present study is set out in the following paragraphs.

Developing a Model to Predict Maintenance Errors

Regarding the relations between safety climate and maintenance errors, there is now a substantial body of empirical evidence from the general safety literature to support the contention that measures of climate are related to safety outcomes. This relationship has been demonstrated in cross-sectional surveys

where scores on safety climate scales have been linked with accidents (Donald & Canter, 1994; Zohar, 1980), in longitudinal studies (Neal & Griffin, 2002), in intervention studies (Donald & Young, 1996), in individual as well as group-level studies (Hofmann & Stetzer, 1996; Zohar, 2000), and across a very wide range of industrial settings. These settings include hospitals (Neal, Griffin, & Hart, 2000), the offshore oil industry (Mearns, Flin, Gordon, & Fleming, 2001), the power industry (Donald & Young, 1996), and chemical processing plants (Hofmann & Stetzer, 1996).

Most of these studies used regression and bivariate correlations to demonstrate the existence of a relationship between safety climate and safety performance. However, a small group of studies have used path analysis or structural equation modelling (SEM) to explain the observed relationships (e.g., Hofmann & Morgeson, 1999, Neal et al., 2000, Tomás, Melia, & Oliver, 1999; Oliver, Cheyne, Tomas, & Cox, 2002). Together, the two groups of studies provided the basis for a hypothesized SEM model that was expected to capture variance in self-reported maintenance errors. The model is shown in Figure 1.
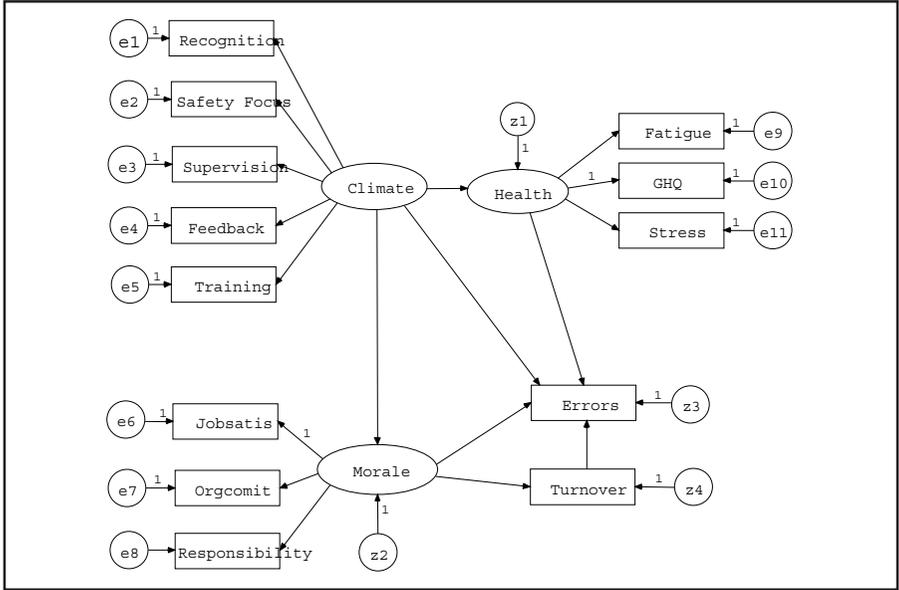
Figure 1. Hypothesised model showing relations among Climate, Morale, Health, Turnover, and Errors.

It can be seen that the full model contains both a measurement and a structural component. Description of the variables that make up the measurement component is deferred until the Method section. A brief description of the structural component is presented here. The first component of the model concerns the safety climate section of the MES. James and James (1989) argued that the various dimensions of climate reflect a higher-order factor (General Psychological Climate, PC_g). Safety climate variables are therefore shown as indicators of

a latent Safety Climate construct (James & James, 1989). Safety Climate was expected to influence a second latent construct labeled Morale, which was measured by the Commitment, Job Satisfaction, and Responsibility variables. Safety Climate also was expected to influence the psychological health of the individual maintenance workers; a construct that has been labeled simply as Health and that was captured by the measures of stress, fatigue, and the General Health Questionnaire (GHQ: Goldberg and Williams, 1988). Support for these separate pathways to Morale and Health can be found in the work of Hart (1994) showing that morale and psychological distress are separate outcomes of positive and negative work experiences. Support for the mediating role of psychological health can be found in the work of Oliver et al. (2002) who examined the relationships between individual psychological, work environment, and organizational variables and occupational accidents using SEM. They found that the individual level variables, including safe behaviour and general health, mediated the effects of the organizational variables on accidents. Stress, in particular, was an important mediator of both organizational and environmental variables.

The pathway from Morale to Turnover was based on well-replicated organizational research demonstrating a strong inverse relationship between commitment and job satisfaction, on the one hand, and turnover intentions on the other (Hulin, 1991). All three latent constructs – Climate, Morale, and Health – were expected to contribute to the variance in self-reported errors. From a theoretical point of view, the role of turnover intentions and its relationship with errors was not very clear. Reflecting the exploratory nature of some aspects of this study, and bearing in mind its expected relationship with Morale, Turnover is shown as influencing Errors.

Method

Participants

A total of 240 maintenance engineers (232 males) working at the two main helicopter repair bases for the Australian Army responded to the survey, representing a response rate of over 90%. Supervisors, inspectors, and higher-level managers also were surveyed but their responses will not be considered here. The survey was targeted primarily at tradespersons (79%) and trainees (21%). The average age of the respondents was 28.5 years and most respondents (84%) had been working as a maintenance engineer or a trainee engineer for at least one year.

Materials

In many instances, scales were already available to measure particular constructs of interest to this study but the approach taken here was that the questionnaire should be tailored to a maintenance context. Accordingly, although individual items may be the same as those used by other researchers, each of

the scales was developed for the purpose of this study. The questionnaire commenced with a series of 12 demographic questions relating to age, training, years of service, and particular area of expertise (e.g., avionics). Unless otherwise indicated, the remaining questions in the survey employed a five-point Likert scale format where 1 indicated strong disagreement and 5 strong agreement. Some items were reverse-scored to encourage respondents to read each question carefully. Scales were formed on the basis of the factor analysis and average scores obtained by dividing total scores by the number of items in the scale. The scales are described below.

A. Safety Climate

1. Recognition for doing good work (5 items). Sample item: In this job, people are rewarded according to performance.
2. Safety focus of the organization (5 items). Sample item: This unit regards safety as a major factor in achieving its goals.
3. Supervision standards (6 items). Sample item: My immediate supervisor really understands the maintenance task.
4. Feedback on work performance (7 items). Sample item: The quality of our work is rated or evaluated frequently.
5. Training standards and appropriateness (5 items). Sample item: My training and experience have prepared me well for the duties of my current job.

B. Morale

6. Job satisfaction (7 items). Sample item: I like maintenance work.
7. Commitment to the organization (7 items). Sample item: I am proud to tell others that I am part of this unit.
8. Sense of personal responsibility (5 items). Sample item: Whether or not my job gets done is clearly up to me.

C. Psychological Health

9. Exposure to workplace stressors (10 items). Sample item: I get anxious when I work to strict deadlines.
10. Fatigue (5 items). Sample item: My overall sleep quality is extremely poor.
11. Psychological Health. The abbreviated, 12-item form of the General Health Questionnaire (GHQ: Goldberg and Williams, 1988) was used. The GHQ explores four aspects of psychological health: somatic symptoms; anxiety and insomnia; social dysfunction; and severe depression. High scores indicate poor psychological health.

D. Outcome Variables

12. Turnover intentions (1 item). Respondents were required to indicate one of three options: whether they intended to keep working in the maintenance industry (scored 1), whether they were uncertain (2), or whether they were resolved to leave the industry (3).
13. Maintenance errors (5 items). Sample item: I make errors in my job from time to time.

E. Affectivity

Positive (PA) and negative affectivity (NA) were measured using the Positive

and Negative Affectivity Schedule (PANAS: Watson, Clark, & Tellegen, 1988). The schedule consists of 10 positive and 10 negative adjectives that respondents rate on a 5-point Likert scale, in terms of how they have felt over the last six weeks. High scores on each scale denote higher levels of affectivity. Watson et al. (1988) reported internal consistency reliabilities for PA and NA of .87 and .88 respectively. Eight week test-retest reliabilities were .68 for PA and .71 for NA.

Procedure

The survey was sponsored by Army Aviation Headquarters and survey forms were included in the pay envelopes of all maintenance personnel along with a covering letter explaining the purposes of the survey. To ensure anonymity, self-addressed envelopes were included so that the forms could be returned directly to the investigator. At the completion of the study, feedback sessions on the main findings of the study were conducted by the investigator and a research assistant.

Results

After initial data screening with SPSS (version 11.0.1) to check for accuracy of data entry, the first stage of the analysis involved the reduction of the 112 items comprising MES to a manageable set of underlying factors. The maximum likelihood method of exploratory factor analysis (EFA) with oblique rotation was used for this purpose. Thirty-two of the 112 items came from well-validated scales (e.g., PANAS and GHQ) and a further 12 items were concerned with demographic data, so these were not subjected to factor analysis. The remaining 68 items were developed or adapted for the purposes of the present study and formed too large a block to factor analyze simultaneously. Accordingly, a strategy was adopted wherein groups of items that were intended to measure a particular construct (Climate, Morale, or Health) were factor analyzed separately. Where there was evidence of unidimensionality and where reliability analysis suggested that a scale formed from the items had good internal consistency (Cronbach's alpha), the construct was retained for further analysis.

As a result of these analyses, six items and the Responsibility scale (5 items) were discarded. All remaining scales, except for the error scale, had satisfactory reliability estimates with alpha estimates above .70. The low reliability of the error scale (.60) was of concern but was still adequate for research purposes (Nunnally & Bernstein, 1994). Means, standard deviations, and reliability estimates for the scales are shown in Table 1 to provide background information about sample characteristics.

Table 1
Summary Statistics for MES Scales (N = 240)

Scale	Number of Items	Mean	SD	Alpha
Recognition	5	2.60	.73	.81
Safety Focus	3	3.81	.63	.76
Supervision	6	3.46	.61	.84
Feedback	7	2.99	.48	.72
Training	4	3.49	.63	.76
Satisfaction	7	3.38	.59	.86
Orgcommit	7	3.08	.70	.86
Stress	10	3.08	.49	.77
G H Q	12	1.88	.45	.89
Fatigue	4	2.56	.73	.79
PA	10	3.03	.79	.91
N A	10	1.61	.57	.84
Turnover	1	1.53	.75	
Errors	3	3.33	.61	.60

Note. Reliability could not be estimated for the Turnover scale

The third column in Table 1 shows the average rating of all respondents on a scale ranging from 1 to 5 (except for the GHQ, where scores ranges from 0 to 4). Scores for many scales were reflected so that - with the exception of Stress, Fatigue, GHQ, Turnover, NA, and Errors - a high score is desirable. We can see from these statistics that this sample could be described as having a high concern for safety; as being well-supported, well-trained, and well-supervised; as being satisfied with their jobs but desirous of more recognition; as having moderate levels of fatigue, low levels of negative affectivity, and as being prepared to admit to making job-related errors. These statistics are in keeping with a military maintenance organization most of whose members, at the time of the survey, were engaged in normal work schedules in peacetime conditions. Another feature of these data is that the standard deviations indicate a reasonable spread of scores on all variables, suggesting that there were individuals in the data set who could not be characterized by the above description. This variation in response patterns forms the basis for the analysis of relations among all variables, leading ultimately to the test of a structural model that links the variables in a causal network. The correlations will be examined first; they are shown in Table 2.

Table 2
Correlations Among Variables (N = 240)

Scales	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Recognition	1.00												
2. Safety Focus	.31												
3. Supervision	.27	.30											
4. Feedback	.44	.28	.39										
5. Training	.42	.37	.38	.43									
6. Satisfaction	.36	.23	.21	.18	.44								
7. Orgcommit	.45	.34	.38	.30	.42	.48							
8. Stress	-.33	-.17	-.15	-.23	-.29	-.20	-.18						
9. GHQ	-.35	-.21	-.18	-.38	-.41	-.35	-.27	.44					
10. Fatigue	-.14	-.04	-.07	-.14	-.17	-.11	-.09	.14	.20				
11. PA	.36	.23	.26	.34	.42	.51	.48	-.27	-.51	-.19			
12. NA	-.22	-.10	-.23	-.22	-.20	-.17	-.14	.42	.70	.15	-.23		
13. Turnover	-.24	-.07	.00	-.09	-.14	-.30	-.33	.13	.20	.07	-.22	.15	
14. Errors	-.16	-.12	-.09	-.13	-.19	-.07	-.10	.48	.31	-.05	-.15	.25	.04

Note. Correlations above .15 are significant at .01 level

The main dependent variable, Errors, is shown on the bottom line. It can be seen that five of the independent variables were significantly correlated ($p < .01$) with Errors. The highest correlation was with Stress, which on its own accounted for 23% of the variance in self-reported errors. Other variables with significant ($p < .01$) associations included GHQ ($r = .31$), NA ($r = .25$), Training ($r = -.19$), and Recognition ($r = -.16$).

Modelling the Interactions Among Organizational and Individual Variables and Errors

The main aim of the present study was to build a model that captures the major sources of variance in maintenance errors. Before attempting this step, however, it was first necessary to deal with the potential criticism that observed relations among the climate measures in Table 2 could simply be reflecting biases that are inherent in self-report measures (Danna & Griffin, 1999; James & James, 1989). There are many ways to deal with method variance, none of them completely effective. NA and PA are often used as direct measures of a tendency to respond in a positive or a negative way to self-report items. The influence of these variables can then be removed statistically, leaving the partial correlations relatively free of method variance. This technique has its drawbacks (see Podsakoff, Mackenzie, Lee, & Podsakoff, 2003 for a review), the main criticism relating to the fact that NA and PA capture more than method variance and that partialling out their influence may lead to serious underestimations of the strength of relations among variables. However, objections about method variance can be overcome if it can be shown that substantial variance still remains after NA and PA have been controlled using hierarchical regression analyses. That was the case in the present study: NA and PA accounted for a significant 6% of the variance in errors when they were entered as the first step in a hierarchical regression analysis. However, as will soon be demonstrated, this represented a small part of the overall variance captured by the structural model. Furthermore, when all variables were entered in the regression equation, the contribution of PA and NA was not significant, suggesting that method variance was not a problem.

Structural equation modelling (SEM), using Version 4.0 of Arbuckle's (1999) AMOS program, was then employed to test the hypothesized model of the relations among the MES variables. Because of the unfavourable ratio of free parameters to cases, a partially aggregated model (Gribbons & Hocevar, 1998) was used wherein subscales based on the EFA represented the various first-order constructs in the conceptual model. The choice of fit indices in SEM is often a controversial matter. In this study, three indices of model fit were used. The first index was the ratio of c^2 to degrees of freedom where Kline (1998) proposed that a ratio of less than three is acceptable. One incremental fit index was used; the comparative fit index (CFI: Bentler, 1990), which is considered to be reasonably robust against violations of assumptions and where a value above .90 was considered to indicate satisfactory fit. The third index used was the root mean square error of approximation (RMSEA: Steiger, 1990), which indicates

the mean discrepancy between the observed covariances and those implied by the model per degree of freedom, and therefore has the advantage of being sensitive to model complexity. A value of .05 or lower indicates a good fit and values up to .08 indicate an acceptable fit (Kline, 1998).

A test of the full structural model shown in Figure 1 (with Responsibility removed) yielded acceptable fit indices (CMIN/DF = 1.85; CFI = .93; RMSEA = .06). The model predicted 47% of the variance in (psychological) Health, 65% of the variance in Morale, 27% of the variance in Turnover, and 39% of the variance in Errors. However, none of the direct paths from Climate, Turnover, and Morale to Errors were significant. Following a strategy of deleting each of these pathways in turn and using the Chi Square difference test to note the effect on fit indices, it was found that the only the last of these three pathways (Morale to Errors) was needed to maintain good fit.

Given the exploratory nature of this research, modification indices were inspected to check the possibility that other theoretically justifiable changes may improve the model. The only noteworthy suggestion involved the fitting of a pathway from Fatigue to Errors. In other words, Fatigue shared variance with Errors that was not captured by the latent trait, Health. Further investigation of this part of the measurement model showed that although Fatigue loaded on Health, the loading was not strong and there was justification for redefining it as a stand-alone variable. This change had a flow-on effect and several other pathways in this part of the model were revised. The final model, with parameter estimates, is shown in Figure 2. This model fitted the data (CMIN/DF = 1.68; CFI = .95; RMSEA = .05) and all pathways were significant.
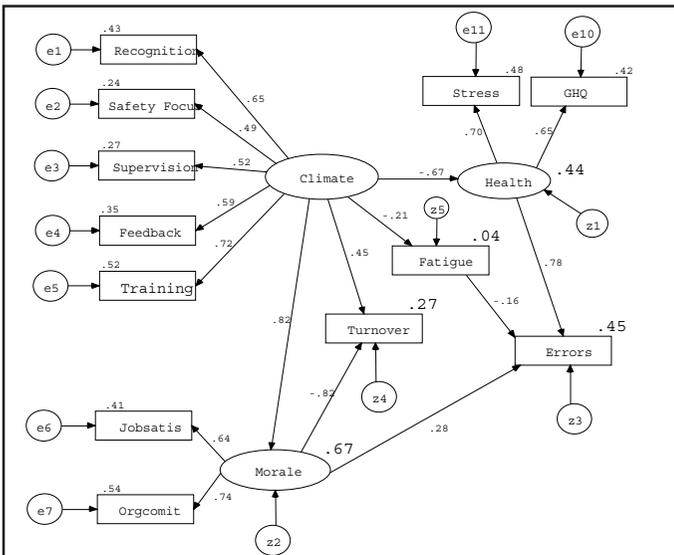

Figure 2. Final model depicting interactions among Climate, Morale, Strain, Fatigue, Turnover, and Errors

For readers not familiar with SEM diagrams, the model can be interpreted as follows. Climate is a latent variable, measured by five different scales. The arrows (pathways) branching from Climate to Health, Fatigue, Turnover, and Morale indicate that Climate is hypothesized as influencing these other four variables. The figures on each of the pathways are standardized path coefficients, which can vary between plus and minus one. A high positive value for a path coefficient indicates that increases in the variable at the start of the path are associated with increases in the variable at the end. A high negative coefficient indicates that increases in the first variable are associated with decreases in the second. A coefficient close to zero indicates that the two variables are not related and that there is no justification for having a pathway linking them.

To illustrate further, Health is measured by two variables, Stress and GHQ. The pathway from Climate to Health has a coefficient of -.67, indicating that better psychological climate leads to better psychological health (because of the way in which the markers were scored, a low score on Health was desirable). Climate has another pathway leading to Fatigue. The negative coefficient (-.21) indicates that as Climate improves, Fatigue decreases. A third pathway from Climate leads to Turnover. In this case, the positive coefficient (.45) does not make sense from a theoretical viewpoint because there is no reason why a more favorable climate would lead to higher job turnover. A check of the correlations between the individual climate measures and Turnover (Table 2) shows that the true relationship is negative: indicating higher job turnover when the climate is poor. Reversals of sign in path coefficients can occur when predictors of a dependent variable are themselves correlated. In the present case, both Morale and Climate are used to predict Turnover and these two predictors are highly correlated. Delete the pathway from Morale to Turnover and the coefficient for the pathway linking Climate and Turnover switches to a negative sign, as one would expect on a theoretical basis. The final pathway from Climate is that leading to Morale and it can be seen that there is a strong positive association between these variables (β = .82). Morale, in turn, has a strong negative relationship (β = -.82) with Turnover. In other words, the better the morale, the less likelihood there is that workers will think about leaving the organization.

There are three pathways leading to Errors. The first of these is from Morale (β = .28). Again, the direction of the relationship is not in the expected direction because of the presence of correlations among the predictor set. Deleting the pathway from Strain to Errors results in the pathway from Morale to Errors switching to its true negative sign, indicating that high morale is indeed associated with lower error rates, as one would expect on an *a priori* basis. The major predictor of Errors in this model is Health which, when tested on its own in a reduced model (not shown here), captured 30% of the variance.

The main features of the model are the impressive R^2 values for all dependent variables. Safety Climate accounted for 44% of the variance in Health, 67% of the variance in Morale, and a small 4% of the variance in Fatigue. Together with

Morale, it also accounted for 27% of the variance in job turnover intentions. Together, these variables accounted for 45% of the variance in self-reported errors. The demonstration that safety climate measures can be modeled using the hierarchical arrangement shown in Figure 1 supports other researchers who have argued for a hierarchical model of climate (e.g., James & James, 1989; Parker et al., 2003).

Discussion

The Maintenance Environment Survey (MES) provided two sorts of data: descriptive data and data pertaining to relations among variables thought to be important in maintenance. Both types of data have proved valuable in this quest to uncover precursors to maintenance errors. The descriptive data, collected from 240 respondents, paints a picture of overall satisfaction with many aspects of the workplace. There was strong support for the level of training and the quality of supervision, two areas that were often criticised during the interviews. Thus, the survey proved a useful counterbalance to some impressions acquired through the interviews.

At the survey level, MES captured some of the major factors relating to work performance. The model (shown in Figure 2) helps to explain seemingly conflicting findings emerging from the analysis of the incident database and the interview data. As mentioned before, the incident reports tended to put the spotlight on human error as the cause of incidents. This is not a surprising outcome; Shappell and Wiegmann (1997) noted that such reporting systems generally focus on identifying "human failures without regard for why the failures occurred" (p. 270). Figure 2 shows that the causal path is as follows: organizational factors influence individuals, who in turn make the errors. The SEM approach has helped to demonstrate the nature of this link. These findings supported the claims of other researchers who point to the influence that social and organizational factors have on human error (e.g., Patankar, 2002; Reason, 1990; Sutcliffe & Rugg, 1998). The present study extended these findings by demonstrating that these linkages are primarily indirect, mediated by individual differences in psychological health and morale.

Implications for Maintenance Work

Reason (1997) likened the practice of surveying the safety climate of organizations as akin to assessing their safety health. This is a very apt description because it is precisely what is implied by the model presented in this paper. Workers' perceptions of such things as management's commitment to safety, appropriateness of training, availability of resources, and possibly many other variables not measured here, do have links with safety outcomes. Whether the perceptions are justified is irrelevant because the effect of the perceptions is felt on morale and psychological health whether there is justification for the perceptions or not. The model also shows that if morale is affected, workers think about leaving the organization, hardly a desirable outcome given the time and

money already invested in that worker and the time and money that will be expended in recruitment and further training. The demonstration of indirect links between climate and errors (via psychological health and morale) suggests that the mere presence of unfavourable perceptions of organizational factors is not sufficient in itself to lead to errors. Unfavourable organizational conditions place pressure on the individual and when the individual begins to succumb to these pressures, errors begin to occur.

The implications are that we should measure both psychological climate and individual health and morale variables on a regular basis to ensure that there are no problems of this kind developing. Similar suggestions have already been made in relation to the value of attitudinal surveys in the maintenance environment. Baranzini and colleagues described a new training, evaluation, and research tool called The Aircraft Maintenance Attitude Survey (AMAS). The AMAS can be used to improve training effectiveness by focusing on safety relevant characteristics of teams and can also help safety goals by monitoring awareness of human factors variables that are related to safety (Baranzini, Bacchi, & Cacciabue, 2001). The UK Civil Aviation Authority is promoting a similar approach through its Safety Health of Maintenance Engineers (SHoMeO tool; CAA, 2003). A different questionnaire has been used in the present study but the findings provide a strong empirical basis for the use of such surveys. Cox and Cheyne (2000) encouraged the reporting of data gained from such surveys as radar plots. Graphic devices such as star plots can help to monitor the safety climate of the organization and the psychological health of the individuals. They will be especially useful if benchmark comparisons within and across organizations become possible (Mearns, Whitaker, & Flin, 2001).

Limitations of Study

In closing, it is important to recognize the methodological shortcomings of the approach followed in this study. The most evident weakness is the use of a cross-sectional methodology, the weaknesses of which in determining causality are well documented. The use of self-report measures for all variables also is problematic. James and James (1989) raised the possibility that predispositions in affect influence both the general climate factor and the first-order climate factors. In other words, affect could be responsible for the commonality observed among climate measures and also responsible for the correlations between climate and performance. One of the strengths of the present study is that it used PA and NA to capture this type of method variance and, in so doing, demonstrated that substantial correlations exist among all variables even when PA and NA are partialled out.

Another criticism of self-report measures is that they may not correlate with objective measures of performance. In the present context, this criticism would translate into the claim that self-reported errors may not correspond with actual errors in the workplace. This criticism can best be addressed by pointing to various studies that have demonstrated a correlation between safety climate

measures and objective indicators of safety performance (Donald & Canter, 1994; Hofmann & Stetzer, 1996; Zohar, 1980, 2000). Theoretical accounts of the links between attitudes, intentions, and behavior, such as that provided by the Theory of Planned Behavior (Ajzen, 1991), also strongly support the use of self-report measures in safety research (e.g., Fogarty & Shaw, 2003).

A further limitation is that the model tested in the current research program has been fitted to data collected in a military environment. Maintenance engineers working in this setting face some challenges (e.g., demands of military duties) that are not faced by those working in commercial settings. The converse also holds true. The model therefore needs to be tested in different organizational settings. Furthermore, as Fahlbruch and Wilpert (1999) pointed out, with the growing trend towards outsourcing of safety units, it may become necessary to extend the safety climate section of the model to include inter-organizational factors. There is no doubt that these factors are becoming important considerations in the aviation industry where key tasks like maintenance are now routinely conducted by third parties. This is true of military as well as civilian aviation organizations.

To conclude, whilst the FAA understands the implications of the tension that exists between increasing demands for air travel and the economic and logistical forces that put pressure on vital functions such as aviation maintenance, increasing the efficiency of maintenance work is just one approach to the problem. Attempts to increase the efficiency of maintenance work need to consider the dynamics of the work environment as they are perceived and experienced by the maintainers themselves. The model reported in this study represents a mathematical approach to capturing and quantifying these dynamics. The model may lack the compelling concreteness of Reason's (1990) famous Swiss cheese model, but it has the potential to be just as effective in guiding practical interventions designed to improve safety. The model contains branches to other organizational outcomes, such as morale and turnover intentions, which undoubtedly affect the overall efficiency of the organization. Methodologies of the kind outlined in this study can provide an empirical basis for directing and evaluating interventions aimed at improving aviation safety.

References

Ajzen, I., (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes, 50*, 179-211.

Arbuckle, J. L. (1999). *Amos users' guide: Version 4.0*. Chicago: Smallwaters Corporation.

Baranzini, D., Bacchi, M., & Cacciabue, P. C. (2001). A tool for evaluation and identification of training needs in aircraft maintenance teams. *Human Factors and Aerospace Safety, 1*(2), 167-193.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Burke, M. J., Borucki, C. C., & Kaufman, J. D. (2002). Contemporary perspectives on the study of psychological climate: A commentary. *European Journal of Work and Organizational Psychology, 11* (3), 325-340.

Civil Aviation Authority (CAA). (2003). Introduction to the safety health of maintenance engineers (SHoMe) tool. Retrieved January 2004 from http://www.caa.co.uk/docs/33/CAPAP2003_11.pdf.

Courteney, H. (2001). *Safety is no accident*. Paper presented at the Royal Aeronautical Society Conference, London, United Kingdom, 2 May.

Cox, S. J. & Cheyne, A. J. T. (2000). Assessing safety culture in offshore environments. *Safety Science, 34*, 111-129.

Danna, K. & Griffin, R. W. (1999). Health and well-being in the workplace: A review and synthesis of the literature. *Journal of Management, 25* (3), 357-384.

Donald, I. J. & Canter, D. V. (1994). Employee attitudes and safety in the chemical industry. *Journal of Loss Prevention in the Process Industry, 7*, 203-208.

Donald, I. & Young, S. (1996). Managing safety: An attitudinal-based approach to improving safety in organizations. *Leadership & Organizational Development Journal, 17* (4), 13-20.

Fahlbruch, B. & Wilpert, B. (1999). System safety – an emerging field for I/O psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 14, pp. 55-93). Chichester: Wiley.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*, 177-192.

Fogarty, G. & Shaw, A. (2003). *Safety climate and the Theory of Planned Behavior: Towards the prediction of unsafe behavior*. Proceedings of the 5th Australian Industrial & Organizational Psychology Conference, 26-28 June, Melbourne, Australia.

Goldberg, D. & Williams, P. (1988). *A User's guide to the General Health Questionnaire*. Windsor, Berkshire: NFER-Nelson.

Gribbons, B. G. & Hocevar, D. (1998). Levels of aggregation in higher level confirmatory factor analysis: Application for academic self-concept. *Structural Equation Modelling, 5* (4), 377-390.

Griffin, M. A. & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety behavior, knowledge, and motivation. *Journal of Occupational Health Psychology, 5* (3), 347-358.

Guldenmund, F. W. (2000). The nature of safety culture: A review of theory and research. *Safety Science, 34*, 215-257.

Hart, P. (1994). Teacher quality of work life: Integrating work experiences, psychological distress and morale. *Journal of Occupational and Organizational Psychology, 67*, 109-132.

Hofmann, D. A. & Morgeson, F.P. (1999). Safety-related behaviour as a social exchange: The role of perceived organizational support and leader-member exchange. *Journal of Applied Psychology, 84*, 286-296.

Hofmann, D. & Stetzer,A. (1996). A cross-level investigation of factors influencing unsafe behaviours and accidents. *Personnel Psychology, 49* (2), 307-339.

Hulin, C. L. (1991). Adaptation, persistence, and commitment in organization. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 445-505). Palo Alto, CA: Consulting Psychologists Press.

ICAO (2003). Human factor guidelines for aircraft maintenance manual.

James, L. A. & James, L. R. (1989). Integrating work environment perceptions: Explorations into the measurement of meaning. *Journal of Applied Psychology, 74* (5), 739-751.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Mearns, K., Flin, R., Gordon, R., & Fleming, M. (2001). Human and organizational factors in offshore safety. *Work & Stress, 15* (2), 144-160.

Mearns, K., Whitaker, S. M., & Flin, R. (2001). Benchmarking safey climate in hazardous environments: A longitudinal, interorganizational approach. *Risk Analysis, 21* (4), 771-786.

Mearns, K. J. & Flin, R. (1999). Assessing the state of organizational safety—Culture or climate? *Current Psychology: Developmental, Learning, Personality, Social, 18* (1), 5-17.

McKenna, J. T. (2002). Maintenance resource management programs provide tools for reducing human error. *Flight Safety Foundation Flight Safety Digest,* 1-15.

Neal, A. & Griffin, M. A. (2002). Safety climate and safety behaviour. *Australian journal of Management, 27* (Special Issue), 67-75.

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety Science, 34*, 99-109.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Oliver,A., Cheyne, A, Tomás, J. M., & Cox, S. (2002). The effects of organizational and individual factors on occupational accidents. *Journal of Occupational and Organizational Psychology, 75*, 473-488.

Parker, C. P., Baltes, B. B., Young, S. A., Huff, J. W, Altmann, R. A., Lacost, H. A., & Roberts, J. E. (2003). Relationalships between psychological climate perceptions and work outcomes: A meta-analytic review. *Journal of Organizational Behavior, 24*, 389-416.

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

Patankar, M. S. (2002). Causal-comparative analysis of self-reported and FAA rule violation data dataset among aircraft mechanics. *International Journal of Applied Aviation Studies, 2* (2), 87-100.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88* (5), 879-903.

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

Reason, J. & Hobbs, A. (2003). *Managing maintenance error.* Brookfield: Ashgate.

Senders, J. W. & Moray, N. P. (1991). *Human error: Cause, prediction, and reduction.* Hillsdale, NJ: Lawrence Erlbaum.

Shappell, S. A. & Wiegmann, D. A. (1997). A human error approach to accident investigation: the taxonomy of unsafe operations. *International Journal of Aviation Psychology, 7* (4), 269-291.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Sutcliffe, A. & Rugg, G. (1998). A taxonomy of error types for failure analysis and risk assessment. *International Journal of Human-Computer Interaction, 10* (4), 381-405.

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

Tomás, J. M., Melia, J. L., & Oliver, A. (1999). A cross-validation of a structural equation model of accidents: Organizational and psychological variables as predictors of work safety. *Work and Stress, 13*, 49-58.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1062-1070.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (1999). Organising for high reliability: Processes of collective mindfulness. *Research in Organizational Behavior, 21*, 81-123.

Wiegmann, D.A., von Thaden, T. L., Mitchell, A. A., Sharma, G., & Zhang, H. (2003). *Development and initial validation of a safety culture survey for commercial aviation*. (Tech. Report AHFD-03-3/FAA-03-1). Federal Aviation Administration, Atlantic City International Airport, NJ.

Zohar, D. (1980). Safety climate in industrial organizations: theoretical and applied implications. *Journal of Applied Psychology, 65*, 96-102.

Zohar, D. (2000). A group-level model of safety climate: testing the effect of group climate on micro-accidents in manufacturing jobs. *Journal of Applied Psychology, 85*, 587-596.

Task Demand and Response Error in a Simulated Air Traffic Control Task: Implications for Ab Initio Training

L. L. Murphy
Department of Psychology, Institute of Simulation and Training,
University of Central Florida

K. Smith
Department of Mechanical Engineering,
Linköping Institute of Technology

and

P.A. Hancock
Department of Psychology, Institute of Simulation and Training,
University of Central Florida

Abstract

This study investigated the relationship between task demand and the occurrence of error in an experimental simulation, which represented dynamic En-Route air traffic control. Participants were trained to baseline performance in the air-traffic-control task and then were presented with a series of 12 challenging but realistic scenarios. These scenarios were scripted to create two cycles of three levels of task demand as represented by traffic count. Conflict opportunities were scripted into each level of traffic count of six conflicts per scenario. Errors of omission were found to be equally likely when traffic count decreased from a peak as during a peak itself. This empirical finding was consistent with real-world experiences as reported in testimonial accounts by professional air traffic controllers. Given the restricted number of participants and their relative inexperience, we consider the present work to be an initial window into a highly complex issue. Our present findings have implications for ab initio training of air traffic controllers and also relate to performance in all operational domains, which demand flawless response from process control agents.

Introduction

The evolving complexity of human-machine systems has served to increase the demands placed upon its operator's limited capacity to process information (see Rochester & Komos, 1976; Smolensky & Stein, 1998). The fluctuating profile of task demand posed in dynamic environments can be expected to elicit a spectrum of behavioral, subjective, and physiological responses, which contribute to operator workload and performance (see Hancock, 1997; Melton, 1982). The increasing trend toward ever more complex technologies, which tax the human information processing system, makes it crucial to develop a thorough understanding of the relationship between task demand, the operator's response to that demand, and the subsequent outcome reflected in the on-going level of performance efficiency.The present study investigated these relationships through the specific and chosen use of a simulated air traffic control (ATC) environment, which permitted the careful quantitative manipulation of task demand level through control of the number of aircraft to be monitored. Since air traffic control represents a dynamic process in which there are explicitly defined operational errors (Greene, Muir, James, Gradwell & Greene, 1997; Hancock, 1997; Hopkin, 1995; Metzger & Parasuraman, 2001), it offers the opportunity to collect a wealth of response data making it an ideal setting for the empirical investigation of task demand and cognitive workload in a success-critical environment.

One of the significant and continuing problems in human performance assessment concerns the meaning of the term *workload*. Some researchers use the term workload to represent objective characteristics of externally based tasks, while others use the same term to represent the subjective experience of the individual placed in that environment (cf., Hancock & Meshkati, 1988; Stager & Hameluck, 1990; Stager, Hameluck & Jublis, 1989). It is evident that this usage as a property of both the operator and the operator's environment has led to a debilitating confusion. We here distinguish between workload as task demand (the external properties of the task independent of any individual) and the operator's response to that task demand, which is assuredly contingent upon the capabilities and skills of the individual so exposed. We suggest that the construct of workload is best defined as a subjective subset of an individual's response to task demand, being closely aligned with the notion of adaptation advanced by Hancock and Warm (1989). Given our representation of workload as a response of the exposed individual, we define *task demand* as a property of the environment itself. Clearly, in this division, the operator's workload response is contingent upon his or her appraisal of any immediate environment as presenting a *task* in the first place. Where such appraisal does not result in the perception of an immediate task to be resolved, no associated workload is experienced (Hancock & Chignell, 1988). Often this specification of a *task* is a result of a third-party arbiter who dictates exactly what that *task* is (Smith & Hancock, 1995). In this present experiment, we, the experimental team, imposed the task. In the real world, tasks arise from a myriad of necessities. In ATC, the controllers' task derives from the ultimate need to provide safe and efficient air transportation.

In a constantly changing task environment, there are often many different sources of information. This profile of information is rarely exactly repeated and for each situation is arguably unique (Hancock, Flach, Caird & Vicente, 1995; Hancock & Meshkati, 1988). Task demand is crucially dependent on the type and amount of task-relevant information presented, and information changes the operator's behavior. We illustrate this by examining the task demand in simulated air traffic control where operators seek to maintain at least the minimum legal separation between aircraft (see Smith, Scallen, Knecht & Hancock, 1998). The level of task demand is a complex function of not only the assimilation and evaluation of task-relevant information, but also of the need to execute a number of cognitive and physical behaviors. The cognitive behaviors include recognizing sources of information on both analog and alphanumeric displays and deciding to intervene (or not) by issuing commands to the pilot (Smith & Murphy, 2000; Smolensky & Stein, 1998). The physical behaviors include monitoring the display, verbally issuing commands to the pilot, handing off aircraft to adjacent sectors, and organizing flight strips (Hopkin, 1995). Sources of task-relevant information include display symbols for aircraft, their vectors and intent, flight strip identifiers for aircraft and their flight-plans, and verbal and written input from traffic management coordinators, from the national weather service, from adjacent sector controllers, and from pilots themselves (Smith & Mafera, 2000). In our simulation-experiment, all sources of information other than traffic count were held constant. The independent variable in the experimental scenarios - the increasing, peaking, and decreasing of traffic count – was therefore the only manipulation of task demand in our study.

We define *response to task demand* as the ensemble of mental actions, overt behaviors, and physiological responses that follow from the operator's interaction with his or her task environment in order to fulfill the goal of the task (Hancock & Desmond, 2001; Melton, McKenzie, Polis, Hoffman, & Saldivar, 1973; Melton, McKenzie, Polis, Funkhouser & Lampietro, 1971; Thackray, Bailey, & Touchstone, 1975; Wilson & Corlett, 1999). Both acceptable performance and error are part of the operator's response to task demand. Another natural measure is task-relevant communication. (Carlson, 1982; Chapanis, 1953; Hendy, 1998) Our measures, the occurrence of error and of task-relevant communication, are overt behavioral indicators of the operator's response to the levels of task demand posed by the experiment.

Task Demand and the Occurrence of Errors

Errors are the nemesis of process control agents working in a system like air traffic control that has a high potential for risk (Smith, Briggs, & Hancock, 1997). *Operational errors* occur whenever two aircraft under positive control violate each other's protected zone (a violation of the minimum separation) which is a compound criterion of 5 miles longitudinally and 1000 feet vertically (Rodgers & Nye, 1993). These formal operational errors are distinguished from the *true errors* of omission and/or of commission that necessarily precede them. Operational errors are viewed not as errors *per se* but as the product of true error at some

earlier point in time. Some true errors are errors of omission: the controller failed to take action that would maintain separation. Other true errors are errors of commission: the controller instructed an aircraft to make a maneuver that directly led to an operational error.

Several previous studies have examined controllers' subjective appraisal of workload (light, moderate, or high) and the occurrence of operational errors (Arad, 1964; Kinney, Spahn & Amato, 1977; Redding, 1992; Rodgers & Nye, 1993; Rodgers, Mogford, & Mogford, 1998; Schmidt, 1976; Stager et al., 1989; Stager & Hameluck, 1990). A review of these relevant studies reveals two themes. First, operational errors tend to occur when traffic count is low to moderate (approximately eight aircraft in the sector). Second, when controllers are asked to subjectively rate traffic volume and workload at the time of the operational error, they also tend to rate it as low to moderate.

It is unclear; however, whether these findings reflect (1) a decreased tendency to make operational errors under high traffic conditions or (2) the lower frequency of high traffic conditions in general (Endsley & Rodgers, 1997). Additionally, it is unclear from these reports whether the true error (of omission or of commission or a combination of these two influences) occurred (A) in low and moderate traffic conditions or (B) earlier when the level of traffic was relatively high. In either case, it may have manifested as an operational error only later when traffic had decreased to a low or moderate level. This source of uncertainty in linking the spatio-temporal occurrence in true error to reported operational errors is a major concern, which we have termed *the phase delay dilemma*.

The Phase Delay Dilemma

Phase-delayed errors are common in dynamic environments (Smith et al., 1998). In the ATC domain, the time lag between a true error and the operational error poses a fundamental dilemma: How are we to know when the true error actually occurred? For example, suppose an operational error occurred during a peak in traffic count. Did the true error also occur during this peak or did it occur earlier as the level of traffic was increasing? Due to the phase delay dilemma, previous research into operational errors may not reveal the actual link between true errors and operational errors and between true errors and task demand. It is therefore not reasonable to conclude that an operational error is a function of a particular level of traffic when the relationship between the true errors and the level of task demand still remains uncertain.

Our experiment was designed to address this impasse by scripting air traffic scenarios that controlled for the phase delay dilemma and by obtaining behavioral measures of performance and response to task demand. Our premise is that systematically changing traffic count in realistic En Route air traffic scenarios result in systematic variations in task demand. By creating two peaks of traffic count in a scenario, we control for the potential confound of operator fatigue. By incorporating conflict opportunities within each level of traffic count,

we also address the potential effects of phase-delayed operational errors and make it possible to distinguish between errors of omission and errors of commission.

<center>Experimental Method</center>

Experimental Participants

Three, upper level, undergraduate students volunteered to act as the air traffic controllers over a six-week period. A second trio of students acted as pseudo-pilots. Participants received class credit for their participation. All participants received 6 practice and 12 experimental scenarios over the six-week period. Scenarios were counterbalanced using a selection from a Latin Square. Students were used rather than professional air traffic controllers for two reasons. First, the purpose of the study was to observe and document errors and inexperienced participants are much more likely to make errors than professionals are. Second, recent national security events have made it extremely difficult to work with professional air traffic controllers (Hancock & Hart, 2002), although we would have preferred to do so. We are fully aware that inexperienced participants are inclined to make different errors than experienced participants (Reason, 1990). Therefore, our results may generalize most to novices like those undergoing ab initio training. However, in mitigation of such issues, at the present stage we are more concerned with contributing to basic empirical understanding than with the immediate domain-specific application of such knowledge.

Experimental Task

The student-controllers were responsible for maintaining the Federal Aviation Administration's criterion for minimum separation between aircraft by issuing appropriate verbal commands to the pseudo-pilots. Command options included changing an airplane's heading, altitude, and/or speed. Pseudo-pilots provided verbal confirmation of controller commands and maneuvered their aircraft accordingly. Pseudo-pilots were trained along side the controllers and were instructed to complete the controller's commands as quickly and as accurately as possible. Analysis of scenario histories confirmed their accuracies in responding, which leads us to conclude that this element of the simulation did not influence subsequent results. During the two weeks of the practice session, the students learned how to use the experimental platform. They became adept at monitoring dynamic traffic, at identifying and resolving potential operational errors, and at giving and confirming verbal commands designed to maintain separation. The criterion for baseline performance was resolution of six scripted aircraft conflicts in a 30-minute practice scenario.

Experimental Platform

The experiments were run using the Distributed Air Traffic Information Display Simulator (DATIDS), a full simulation of an ARTCC sector controller's workstation (Klinge, Smith, & Hancock, 1997). The simulator presents a representation of the composite radar screen, the computer read-out display, and the but-

tons and dials used to adjust the settings of the R-side ATC display and an illustration of this simulation is shown in Figure 1.
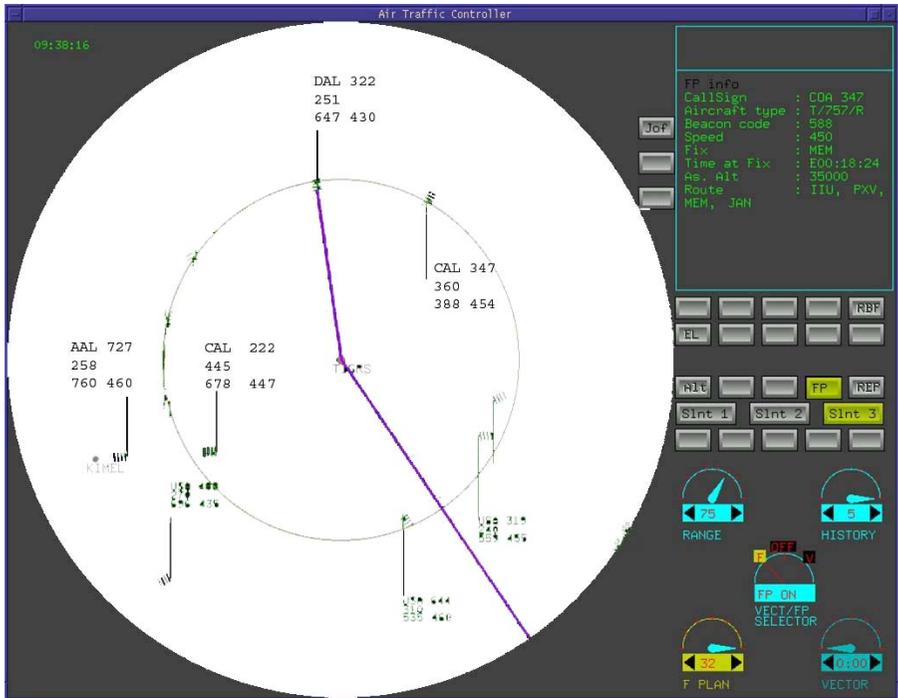
Figure 1. Illustration of the Distributed Air Traffic Information Display Simulator (DATIDS).

Experimental Design

As shown in Figure 2, the experimental scenarios incorporate two blocks of time, each containing a cyclic pattern of traffic count with three levels (5 minutes of increasing traffic, 5 minutes of peak traffic, and 5 minutes of decreasing traffic). The two 15-minute blocks and three levels of traffic count yield a 2x3 repeated measures design. The repeated blocks permit an assessment of the potential confound of fatigue. Task demand is operationally defined as the average traffic count, being the average number of aircraft visible on the controller's information display at each minute. More specifically, in the first period of five minutes of each block, the traffic count continually increased from three aircraft to a peak value of 16 aircraft. During the second period of five minutes of each block, the traffic count remained at or near this peak. In the third period of five minutes of each block, the traffic count gradually decreased until it returned to the baseline of three aircraft. These criteria, of a minimum of three aircraft and a maximum of 16 aircraft, were adopted based on actual observations of a controller's typical sector load (Smith & Murphy, 2000).

Controlling for Phase Delay

To make errors of omission evident, two aircraft were scripted to create two

distinct opportunities for an error of omission in each of the six levels of traffic count. The two aircraft entered and exited the information display during the same (five minute long) level of traffic count. If the student-controller took no action to maneuver these aircraft, then the resulting operational error could be attributed to an error of omission that occurred during that level of traffic and not in a prior level. This control makes it possible to know the number of aircraft and the level of task demand when and if errors of omission occurred. All other aircraft were scripted to be conflict-free. Therefore, if any other errors occurred (at any time) they could be attributed to an error of commission.

Dependent Measures

The primary performance measure was the number of operational errors due to errors of omission or due to errors of commission occurring throughout the experiment. The behavioral measure of response to task demand was the number of commands between controller and the pseudo-pilot.
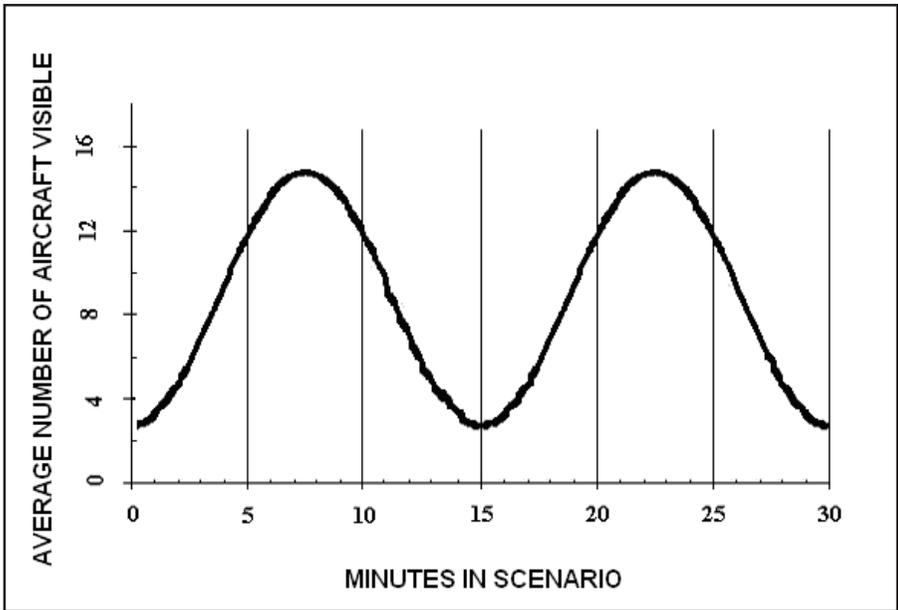
Figure 2. Illustration showing the schematic design of the experimental scenarios.

The horizontal axis is elapsed time in the scenario. The vertical axis is the minute-by-minute average number of aircraft visible on the controller's information display. Each scenario lasted 30 minutes and presented two cycles of traffic count. Each cycle contained three periods of five-minute-long conditions of traffic count: increasing, peak, and decreasing. In each cycle the average number of aircraft increased for a period of five minutes from a baseline of approximately three aircraft, remained near a peak level of approximately 16 aircraft for five minutes, and then decreased for a period of five minutes back to the

base level. Opportunities for operational errors due to errors of omission were scripted into all six periods of five-minute-long conditions of traffic count.

Experimental Results

A repeated measures 2x3 (two blocks of time by three levels of traffic count) analysis of variance was used to analyze the error and communication data. The number of scenarios (12) acted as a proxy for number of participants in order to increase the statistical power of the test. Mauchly's test of sphericity was not significant for any of the analyses and thus sphericity was uniformly assumed.

Errors of Omission

Figure 3 shows the total number of errors of *omission* for each level of traffic count (increasing, peak, and decreasing) and both blocks of time. A 2x3 repeated measures ANOVA was conducted to determine the effect of traffic level and block of time on the number of operational errors due to errors of omission. The main effect for traffic level was found to be significant, $F(2, 22)=10.73$, $p<.001$. A within-subjects $eta^2=.50$ indicates a medium effect size for traffic level. Post-hoc pair-wise comparisons revealed no significant difference between the mean number of errors of omission in the peak and decreasing conditions. However, both the peak and decreasing conditions were significantly different from the increasing condition for both blocks of time, $p<.05$. It appears that the student-controllers were just as likely *not* to take corrective action to resolve potential conflicts after a peak in traffic as they were during a peak in traffic for both blocks of time in the scenario.

Errors of Commission

Figure 4 shows the total number of errors of *commission* for each level of traffic count and both blocks of time. A 2x3 repeated measures ANOVA was conducted to determine the effect of traffic count (increasing, peak, and decreasing) and block of time on errors of commission. Traffic level was not found to be significant. However, there was a significant difference for blocks of time, $F(1, 11)=7.86$, $p<.01$. The within-subjects $eta^2 = .42$ indicates a medium effect size. While many more errors of commission were made in the second block than in the first block, the totals are low but remain significant. Errors of commission were relatively rare. They occurred when student-controllers issued inappropriate commands that caused an operational error between aircraft that otherwise would have remained separated. These results suggest that the student-controllers experienced some amount of fatigue or vigilance decrement (Mackworth, 1948; Mackworth, 1957; Hancock, 1984) or both.
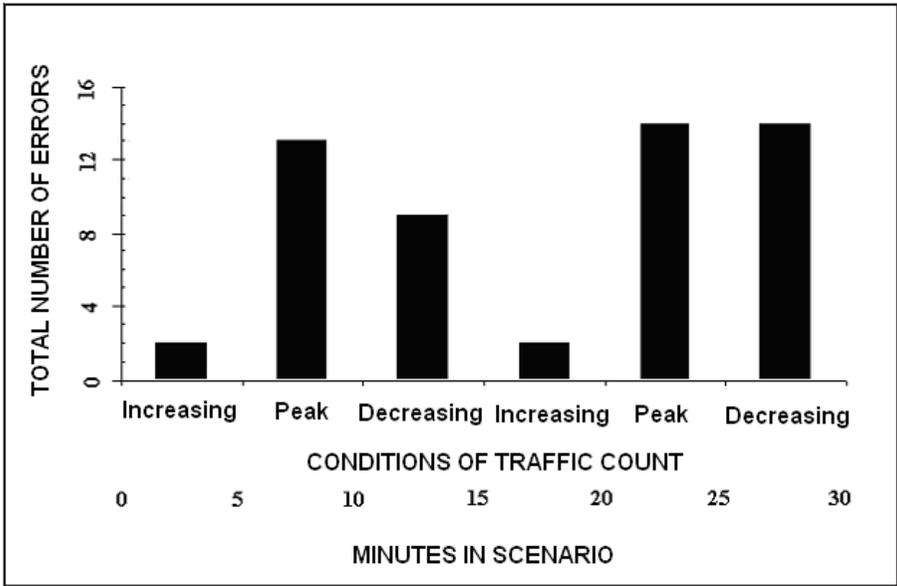
Figure 3. Total number of operational errors due to errors of *omission* collapsed across scenarios and conditions.
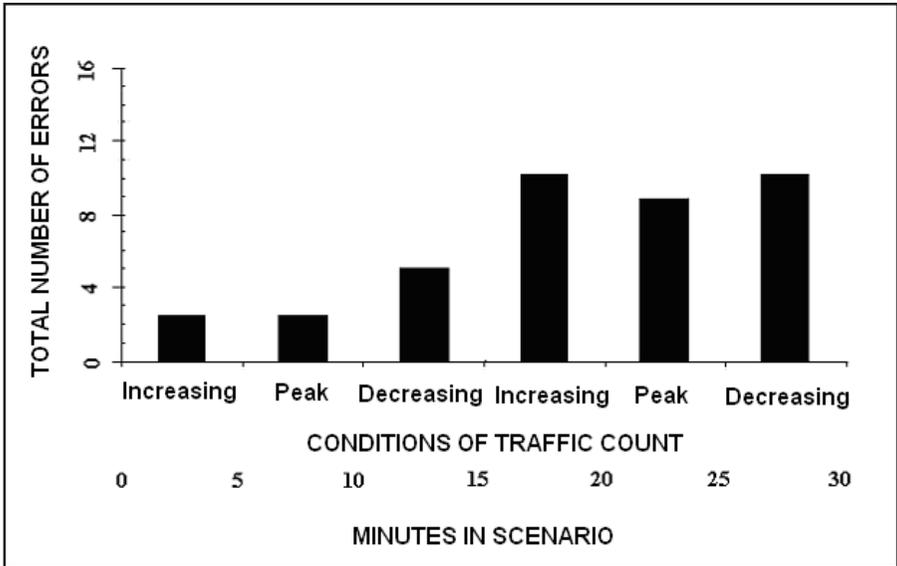
Figure 4. Total number of operational errors due to errors of *commission* collapsed across scenarios and conditions. The vertical scale is the same as in Figure 3 to facilitate comparison.
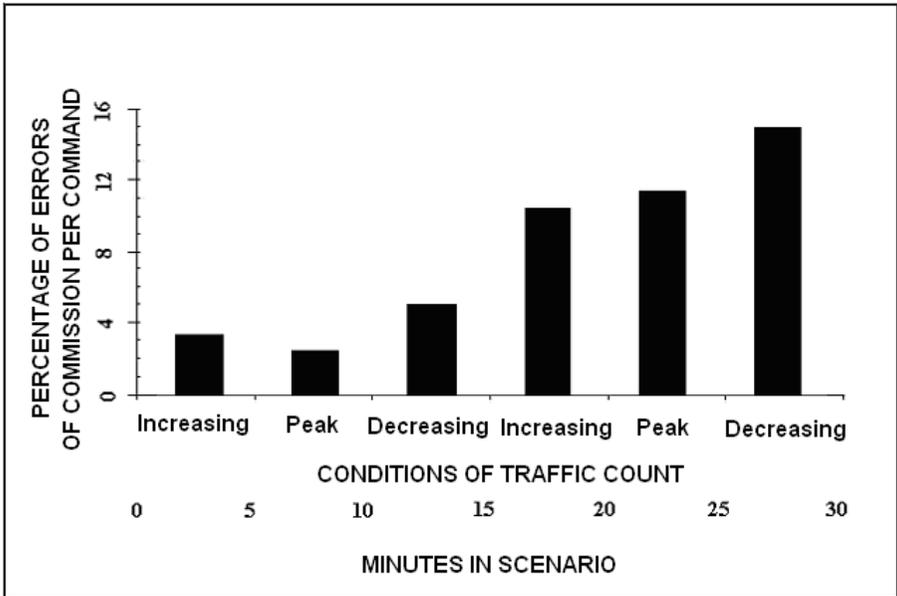
Figure 5. Percentage of errors of commission per command to maneuver aircraft issued by participants collapsed across scenarios and conditions.

Controller Communication

Figure 5 shows the percentage of errors of commission per command to aircraft issued by the student-controllers for each level of traffic count and both blocks of time. A 2x3 repeated measures ANOVA was conducted to determine the effect of traffic level and block of time on the communication. The interaction between the three levels of traffic and the two blocks of time was significant, F (2, 22)=8.29, p<.002. The within-subjects eta^2 = .43 indicates a medium effect size. The interaction suggests that the relationship between communication and errors of commission is such that more errors of commission are committed per command during the second block of time. This may be due to fatigue, diminished attention capacity, or both.

Correlations

A correlation analysis was conducted to detect emergent relationships between the dependent and independent variables. The analysis, shown in Table 1, found significant positive correlations between two of the three dependent variables, errors of omission and communication, and the experimental treatment, level of traffic count. These results indicate that the manipulation of traffic count had an effect on participant behavior. Both types of error were found to correlate significantly with communication. The correlation between communication and errors of commission was positive. This result suggests that errors of commission tended to occur during periods when the student-controllers' response to task demand was relatively high. In contrast, the correlation between communication and errors of omission was negative. This result suggests that

inadvertent operational errors tended to occur when the student-controllers' response to task demand was relatively low.

Table 1

Correlations between the three dependent variables - errors of omission, errors of commission, and communication – and the main dependent variable, traffic count.

	Omission	Commission	Communication	Traffic
Omission	1.00	0.03	- 0.14 *	0.21 **
Commission		1.00	0.22 **	- 0.06
Communication			1.00	0.13 *
Traffic				1.00

* Correlation is significant at the .05 level (2-tailed)
** Correlation is significant at the .01 level (2-tailed)

Discussion

The purpose of this study was to develop an understanding of the relationships between task demand, the operator's behavioral response to that demand, and the occurrence of errors of omission and of commission in a simulated ATC environment. The data suggest that errors of omission are equally likely to occur after a peak in task demand as during that peak and are associated with relatively low levels of response to task demand. Errors of commission are more likely to occur when response to task demand is unusually high and time-on-task has made it likely that fatigue has set in. In short, this type of error is sensitive to fatigue, to the level of task demand, and to the operator's behavioral response to that task demand. Previous investigations of task demand, response to task demand, and the occurrence of errors in the ATC domain (e.g., Kinney et al., 1977; Stager, Hameluck, & Jublis, 1989; Stager & Hameluck, 1990) has focused on workload, the operator's subjective response to task demand. Using archival data, they reported that operational errors were more likely to occur when workload was said to be moderate to low. Since the data were archival, these studies failed to control for the phase delay dilemma, the time lag between the operational error and the error of omission or commission that necessarily preceded them.

Errors of Omission

The total number of errors of omission across conditions shown in Figure 3 revealed that operational errors due to errors of omission were equally likely to occur after a peak in traffic as during a peak. This finding was consistent with anecdotal evidence reported by professional air traffic controllers. Most controllers freely admit that operational errors occur most often on the backside of a

peak in traffic. Unlike previous archival studies investigating operational errors, the current study scripted the simulated traffic scenarios to control for traffic phase-delay. Consequently, the current study contributed a new finding: operational errors occur equally as often on the backside of the peak in traffic as they do during the peak. These results were found for both blocks of time, beginning of scenario (first 15 min) and end of scenario (last 15 min), suggesting that operational errors due to errors of omission are more likely a function of task demand than of fatigue. If the effect were due to fatigue following high traffic levels, then significantly more errors would have occurred at the end of the scenarios, which was not the case. These findings suggested that the decrease of task demand after a peak is just as taxing on inexperienced participants as the peak itself. These participants were expending relatively high levels of effort as they made errors of omission. Operational errors due to omission were not often a result of the inexperienced participant dropping his or her guard.

The correlation between errors of omission and communication was negative in this experiment, see Table 1. Our inexperienced participants issued inappropriate commands to aircraft while missing critical cues about impending conflicts. It appears that the operational errors due to errors of omission in this experiment may reflect our participants' relative lack of skill. Accordingly, these present results may generalize only to inexperienced students in Collegiate Training Initiative institutions that are beginning their training to become air traffic controllers. Whether such results generalize to actual operations or other operations in other process control domains requires further empirical evaluation.

Errors of Commission

The number of errors of commission steadily increased as scenarios progressed, see Figure 4. This result suggested that operational errors due to errors of commission are more a function of fatigue than of the cyclic manipulation of task demand. Additional evidence of a fatigue effect was found in the statistically significant pattern of the errors of commission per command, shown in Figure 5. These findings were congruent with the vigilance literature, which reported that the frequency of correct detections tends to decrease after 20 minutes on watch (e.g., Hancock & Warm, 1989; Mackworth, 1948; Mackworth, 1957; Parasuraman, 1986; Warm, 1984). This deterioration in performance is traditionally termed the "vigilance decrement."

Limitations of the Study and Future Directions

The correlation data in Table 1 show that the cyclic manipulation of traffic was positively correlated with two of the dependent measures, communication and errors of omission. An analysis of variance found that the fluctuation of task demand had an effect on the participants' responses to task demand. These results suggested that this study had high internal validity. However, there are two significant threats to the external validity of this study. The first is nature and the small size of our sample – three participants, all of whom were undergradu-

ate psychology students. However, these students were trained and participated in the study for over six weeks (two weeks of the practice session and four weeks of the experimental session). Therefore, the results may generalize primarily to inexperienced trainees striving to be air traffic controllers. Future studies should of course address a larger group of participants from a pool that is more representative of trainees. That said, it must be noted that Federal and union regulations make it extremely difficult to gain access to these FAA trainees. Eventually if the work is to exert a strong practical impact, it might be replicated and extended with full-time professional controllers. The second threat is construct validity. In debrief sessions none of the student-controllers mentioned the cyclic pattern of task demand and the relatively regular occurrence of the scripted conflicts. Nevertheless, it is possible that they were implicitly aware of them, made tacit hypotheses about them, and responded to those hypotheses. Construct validity may be improved by counterbalancing the order of the presentation of conditions (e.g. from {increase, peak, decrease} to {peak, decrease, increase} etc.), and by adding control conditions with uniform levels of task demand in order to get baselines of performance (see Hancock, Williams, Manning, & Miyake, 1995).

The major finding in this study was that errors of omission are equally as likely on the backside of a peak in task demand as during the peak. It would be useful to determine whether this finding generalizes to other process control and continuous operator tasks and other dynamic environments. Tasks and environments that are amenable to laboratory simulation include operating a motor vehicle, piloting an aircraft, firing weapons, and operating an assembly line. Such studies would be especially effective if the order of the presentation of levels of task demand was fully counterbalanced and control conditions used to get baselines of performance. For example, one useful control condition would pose scenarios with two cycles of traffic demand but no scripted conflicts. Other control conditions would pose scenarios with constant (but different) levels of traffic demand and several scripted conflicts. The resulting within-subject comparisons would provide a strong test of the generality of our results. If this preliminary finding of errors of omission on the backside of peaks in task demand holds up to further testing, it would become imperative to investigate ways of mitigating this effect. Eye tracking during the task could provide detailed information about the information the operator is focusing on (hits vs. noise) and might shed light on these sources of errors and, in turn, on ways these errors might be avoided.

Acknowledgements

References

Arad, B.A. (1964). *Notes on the measurement of control load and sector design in the en-route environment*. (Report No. AD-659035). Washington, D.C.: Federal Aviation Administration.

Carlson, G.E. (1982). *Captain Carlson's airplane talk: The complete book of VFR and IFR communications*. Las Vegas: Watosh Publishing.

Chapanis, A. (1953). *Research techniques in human engineering.* Baltimore: Johns Hopkins Press.

Endsley, M.R. & Rodgers, M.D. (1997). *Distribution of attention, situation, awareness, and workload in passive air traffic control tasks: Implications for operational errors and automation*. (Report No. DAOT/FAA/AM-97/13). Washington, D.C.: Federal Aviation Administration.

Green, R.G., Muir, H., James, M. Gradwell, D. & Greene, L.G. (1997). *Human factors for pilots*. Vermont: Ashgate Publishing Co.

Hancock, P.A. (1984). Environment stressors. In: J.S. Warm (Eds.). *Sustained Attention in Human Performance.* (pp. 103-142), New York: Wiley.

Hancock, P.A. (1997). *Essays on the future of human-machine systems*. MN: Banta Information Services.

Hancock, P.A., & Chignell, M.H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics, 18,* 647-658.

Hancock, P.A. & Desmond, P.A. (2001) (Eds.). *Stress, workload, and fatigue: Human factors in transportation*. Mahwah, New Jersey: Lawrence Erlbaum Associate Publishers.

Hancock, P.A., Flach, J., Caird, J., & Vicente K. (1995). (Eds.). *Local applications of the ecological approach to human-machine systems*. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.

Hancock, P.A. & Hart, S.G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design, 10 (1),* 6-16.

Hancock, P.A. & Meshkati N. (Eds.) (1988). *Human mental workload*. Amsterdam: North-Holland.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519-537.

Hancock, P.A., Williams, G., Miyake, S., & Manning, C.M. (1995). The influence of task demand characteristics on workload and performance. *International Journal of Aviation Psychology, 5,* 63-85.

Hendy, K C. (1998). CRM: More than just talk, talk, talk..*Flight Safety Spotlight, 98(1),* 12-15.

Hopkin V.D. (1995). *Human factors in air traffic control*. England: Burgess Science Press.

Kinney, G.C., Spahn, J., & Amato, R.A. (1977). *The human element in air traffic control: Observations and analyses of the performance of controllers and*

supervisors in providing ATC separation services. (Report No. MTR-7655). McLean, VA: METREK Division of the MITRE Corporation.

Klinge, J., Smith, K., & Hancock, P.A. (1997). DATIDS: The University of Minnesota distributed air traffic information display simulator. *Proceedings of the 9th International Symposium on Aviation Psychology*, Columbus, Ohio.

Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*, 6-21.

Mackworth, N.H. (1957). Some factors affecting vigilance. *Advancements in Science*, *53*, 389-393.

Melton, C.E, McKenzie, J.M., Polis B.D., Funkhouser, G.E & Iampietro, P.F. (1971). *Physiological responses in air traffic control personnel: O'Hare Tower.* (Report No. DOT/FAA/AM-71-2). Washington, D.C.: Federal Aviation Administration.

Melton, C.E, McKenzie, J.M., Polis B.D., Hoffman, M., & Saldivar, J.T. (1973) *Physiological responses in air traffic control personnel: Houston Intercontinental Tower* (Report No. DOT/FAA/AM-93-22). Washington, D.C.: Federal Aviation Administration.

Melton, C. E. (1982). *Physiological stress in air traffic controllers: A review*. (Report No. OAM/FAA/AM-82/17). Washington, D.C.: Federal Aviation Administration.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors, 43(4)*, 519-528.

Parasuraman, R. (1986). Vigilance, monitoring, and search in K. R. Boff L. Kaufman, and J. P.Thomas (Eds.), *Handbook of Perception and Human Performance: Cognitive Processes and Performance* (Vols. 2.). New York: Wiley.

Reason, J.T. (1990). *Human error*. Cambridge, UK: Cambridge University Press.

Redding, R.E. (1992) Analysis of operational errors and workload in air traffic control. *Proceedings Human Factors Society, 36*, 1321-1325.

Rochester, S.I. & Komos, N.A. (1976). *Takeoff at mid century. Federal Aviation policy in the Eisenhower years 1953-1961*. Washington, D.C.: Federal Aviation Administration.

Rodgers, M.D., & Nye, L.G. (1993). Factors associated with the severity of operational errors at air route traffic control centers. In M.D. Rodgers (Ed.), *An examination of the operational error database for air route traffic control centers* (Report No. DOT/FAA/AM-TN93/22, NTIS No. ADA275986). Washington, D.C.: Federal Aviation Administration.

Rodgers, M.D., Mogford, R.H., & Mogford, L.S. (1998). *The relationship of sector characteristics to operational errors* (Report No. DOT/FAA/AM-98/14). Washington, D.C.: Federal Aviation Administration.

Schmidt, D. K. (1976). On modeling ATC workload and sector capacity. *Journal of Aircraft*, *13(7)*, 531-7.

Smith, K., Briggs, A., & Hancock, P.A. (1997). Success and failure at self-separation in simulated free flight. *Proceedings Meeting of the Human Factors and Ergonomics Society*, *41*, 13-17.

Smith, K., & Hancock, P.A. (1995). Situation awareness is adaptive, externally-directed consciousness. *Human Factors*, 37, 137-148.

Smith, K., & Mafera, P. (2000). *ATC coordinator's information requirements for the NAS*. (Report No. AAR-100). Washington DC: The Office of the Chief Scientist and Technical Officer for Human Factors of the Federal Aviation Administration.

Smith, K., & Murphy, L. (2000). *Traffic management unit structure, positions, and uses of the traffic situation display*. (Report No. AAR-100). Washington DC: The Office of the Chief Scientist and Technical Officer for Human Factors of the Federal Aviation Administration.

Smith, K. Scallen, S.F., Knecht, W., & Hancock, P.A. (1998). An index of dynamic density. *Human Factors*, *40*, 69-78.

Smolensky, M.W. & Stein, E. S. (1998). *Human factors in air traffic control*. California: Academic Press.

Stager, P., & Hameluck, D. (1990). Ergonomics in air traffic control. *Ergonomics*, *33*, 436-9.

Stager, P., Hameluck, D., & Jublis, R. (1989). Underlying factors in air traffic control incidents. *Proceedings of the Human Factors Society, 33*, 43-46.

Thackray, R.I., Bailey, J.P., & Touchstone, R.M., (1975). *Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task*. (Report No. DOT/FAA/AM-75-B) Washington, D.C.: Federal Aviation Administration.

Warm, J.S. (Ed.) (1984). *Sustained attention in human performance*. Chichester: Wiley.

Wilson, J.R. & Corlett, E.N. (1999). *Evaluation of human work: A practical ergonomics methodology*, (2nd ed.), Philadelphia: Taylor & Francis Inc.

An Analysis of GPS Usability and Human Performance in Aviation

Humza Adam,
John E. Deaton,
Ronald W. Hansrote,
and
Muzaffar A. Shaikh
Florida Institute of Technology
150 West University Boulevard
Melbourne, FL 32901-6975

Abstract

With the advent of widespread utilization of advanced technological devices in many of the aircraft flying today, it has become absolutely imperative to determine the safety implications resulting from such devices. This study focused specifically on the use of cockpit-mounted GPS receivers. An investigation of GPS receiver interfaces was performed with the ultimate goal of determining appropriate recommendations pertaining to safe, efficient, and effective interface design. Two different popular GPS receiver interfaces (a Garmin 530 and a Bendix/King KLN 89B), as well as pilots with high (Certified Flight Instructor) versus low (Instrument-rated Private or Commercial Pilot) levels of certification, were compared. It was found that, overall, the pilots performed better on several tasks using the graphically-oriented interface (the Garmin 530) than they did on the textually-oriented interface (the Bendix/King KLN 89B). Additionally, the results of a post-experiment questionnaire revealed many potential improvements to existing GPS interface design. The results are discussed and appropriate recommendations offered. Finally, areas of future research are identified.

Introduction

The aviation industry has been one of the greatest benefactors of GPS technology (Wroten, 1999). The increasingly widespread use of this technology is becoming a major driving force in the trend toward enhancing the capacity, efficiency and safety of the National Airspace System (NAS) in the United States and elsewhere around the world (Goel, n.d.). This trend has direct applications

to Free Flight, which aims to provide pilots with greater freedom to determine the routes they shall fly in traveling from one point to another. The greatest advantage of this concept over the current system is that it "moves the airspace concept from a centralized command-and-control system between pilots and air traffic controllers to a distributed system that allows pilots, whenever practical, to choose their own routes and file flight plans that follow the most efficient and economical routes" (Goel, n.d.).

While the future of GPS technology in aviation appears bright, careful consideration must be given to the human performance issues concerning pilots' use of this relatively new technology prior to the implementation of any major changes to the NAS. Indeed, the interactions between humans and their GPS units will, in large part, determine how safe and effective the new system shall be, especially since this system shall rely almost exclusively on the successful interaction between man and machine. Comparatively little research has been done in this area of aviation primarily due to the fact that GPS technology is currently still in its infancy. Thus, in an attempt to help fill this research void, this work aims to provide an understanding of the effects of GPS use on pilot behavior and performance.

Review of Previous Research

To date, there have been relatively few studies that have discussed the impact and effect of GPS use on humans in an aviation context. However, those studies that discuss these issues often do provide a wealth of information on the subject. Specifically, the studies reviewed concerned the following issues: safety concerns over the implementation of the Automatic Dependent Surveillance Broadcast (ADS-B) system for general aviation pilots, and the usability of GPS display screen and menu formats and- procedures. It should be noted that the ADS-B system is a newly proposed air navigation system that will allow aircraft (or obstacles or vehicles) to regularly transmit their GPS derived position and velocity information to other aircraft, as well as ground-based facilities such as Air Traffic Control (ATC). Specifically, these studies relate how automation-related issues, interface-design issues, and human operator issues affect aircraft GPS users (pilots). Each of these issues shall be discussed separately.

Automation-Related Issues

Automation has had tremendous appeal over the past few decades and the newest aircraft cockpits are testament to this fact since they incorporate the great advances in technology to perform tasks that previously could only be performed manually by humans. However, with the proliferation of high-technology automated gadgets comes an increased need to understand potential safety issues arising out of their use. These issues include: complacency, over-reliance and head-down time, to name but a few. Additionally, vigilance decrement issues continue to plague highly automated environments (Parasuraman et al., 1996).

A recent study performed in Alaska to determine how well general aviation pilots would interact with ADS-B displays provides useful insight into the functionality of the future ATC system once the proposed ADS-B system is implemented on a wide scale. In this particular study, about 150 aircraft in Bethel, Alaska were fitted with advanced avionics units required by the ADS-B system (Williams et al., 2002). The displays installed included an Apollo MX-20 multi-function display and an Apollo GX-60 GPS display. The multi-function display was not only capable of providing a moving map display with terrain features, but also weather and traffic information.

The participants were "[US Federal Aviation Regulations] Part 135 airline operators and pilots in the Bethel area" (Williams et al., 2002). After an unspecified period of time of display use, 41 pilots were interviewed and 27 of those pilots also completed self-administered questionnaire forms. The data on the day-to-day use of the displays were collected by a human factors team formed from the FAA's Safe Flight 21 Office and University of Alaska at Anchorage (UAA) personnel.

The data collected indicated certain safety implications. Among these was "degradation of conventional flying skills" (p. 4), which may be a legitimate concern considering the fact that approximately 41% of the pilots believed that their conventional navigational skills had deteriorated as a result of the reliance on the new displays (Williams et al., 2002). Indeed, this can be a serious problem as the over-reliance on such displays can easily lead to a loss of situation awareness upon receiver failure, which can result in fatal accidents such as CFIT (Controlled Flight Into Terrain). Another issue identified by at least one pilot was the fact that certain terrain features, namely "mud volcanoes" (p. 5) were missing from the terrain information displayed. This indicated another instance where over-reliance on the displays with incomplete and/or erroneous terrain information may induce CFIT accidents.

Head-down time is another serious issue of concern when pilots interact with and/or pay unwarranted attention to the displays. Head-down time here may be defined as the time spent tending to the displays at the expense of paying attention to other aircraft instruments and/or visually scanning outside the aircraft for traffic. Williams et al. (2002) noted that increased workload and reduced situation awareness may be some of the consequences associated with the introduction of new equipment into general aviation aircraft, particularly when the equipment has just been installed and no formal training program has been instituted. The researchers found that up to half of the participating pilots had not received formal training on the use of the systems from their (the pilots') flight companies. They also stated that the pilots complained of having to spend a "considerable amount of head-down time attempting to select and exercise system functions" (p.5) during the early flights.

Increased risk-taking behavior due to overconfidence in the displays was another issue identified by the study where up to 83% of the pilots responded that there would be an increased likelihood of flight into low visibility conditions with the displays available (Williams et. al, 2002). This appeared to be an area that raises some serious safety concerns given that Nendick and St. George (1995) also reported that users acknowledged being tempted to fly in similar risky conditions with an available GPS unit. Unfortunately, such behavior can be, and actually has been, a contributing factor to tragic accidents (Heron et al., 1997).

The problems described here such as "degradation of flying skills," head-down time issues, over reliance, and so on, are not simply limited to GPS use only and can be found in almost any aviation environment where a high degree of automation dependency exists. In fact, some of these same problems can be said to have contributed to the infamous 1972 crash of Eastern Flight 401. Flight 401, an L-1011, crashed in the Florida Everglades after the crew failed to detect an "inadvertent autopilot disconnect" while engaged with a possible landing gear malfunction (Wiener, 1988, p. 439). Here, the effects of lengthy crewmember head-down times (due to the distraction) combined with the crew's over-reliance (on the aircraft's autopilot system) to result in a controlled flight into terrain (CFIT) accident.

Interface-Design Issues

Other studies have also focused on the usability of GPS displays; specifically, the ease with which relevant information may be obtained from GPS devices certified for aerial navigation. One such study, wherein an assessment of cockpit GPS menus and procedures was performed, was performed by the Civil Aeromedical Institute (CAMI) (Wreggit & Marsh, 1996). Wreggit and Marsh noted that distraction of pilots from visual scanning and control of aircraft when GPS devices are involved may in fact be the result of poorly designed software interfaces and sub-optimal menu structures. The great and numerous differences amongst GPS manufacturers in design of aviation GPS units and their interfaces are most likely a consequence of non-existent standards for "data entry and retrieval, display type, or placement within the cockpit" (Wreggit & Marsh II, 1996, p.2). This inevitably leads to poor or even negative transference when pilots must use GPS units built by different manufacturers.

In a New Zealand study performed by Nendick and St. George (1995), 172 pilots responded to questionnaires concerning GPS use. Certain findings were similar to those revealed in the study performed by Wreggit et al. (1996). One such finding related to data input errors. Specifically, Nenedick and St. George found that over half of the users sampled (55%) reported input-related difficulties. These input-related difficulties were described to have led to such errors as "hitting the wrong key; forgetting the keying sequence to obtain the correct information; and inadvertently pressing a key twice in turbulence resulting in a change of mode or number or letter" (p. 154). Adams et al. (2001) also specifi-

cally cited input errors among the leading causes of GPS-related problems. Thus, it is not difficult to understand how a poor interface display and design (e.g. poorly spaced keys and buttons, or non-intuitive labels) can induce and exacerbate such errors.

The study performed by Wreggit and Marsh II was conducted with the goal of discovering usability issues brought forth by GPS unit design. This particular study was performed using 9 private pilots considered novice GPS users to avoid negative transfer of training due to prior GPS experience. The participants were required to perform several flight-related GPS tasks such as waypoint setting, general GPS data entry and GPS data retrieval, and GPS navigation while also flying a fixed base simulator termed the Basic General Aviation Research Simulator (BGARS). The GPS unit evaluated was a Magellan EC-10X. The data collected included head-down time, task times, pre-flight and post-flight questionnaire data (Wreggit & Marsh II, 1996). Video recorders were used and pertinent information was captured real-time by the experimenters who took notes throughout the study.

Among the main findings of this study was that a significant correlation existed between head-down time and total number of button presses per task (Wreggit & Marsh II, 1996). Also, excess button presses usually resulted from a misinterpretation of the functions of certain buttons such as "OUT" and "ENT", which were often used interchangeably in error by users to perform the same task such as removal of waypoints. This form of error was termed "double error" (p. 5) where the pressing of "OUT" did not accomplish the task and therefore "ENT" was pressed (erroneously) to perform the same task. Similar excess button presses occurred when attempts to delete certain waypoints were made, but with the incorrect sequence of button presses (such as attempting to delete waypoints while the flight plan was still active).

Another issue that was singled out concerning the particular GPS device used for the study was that error feedback was provided through auditory messages in the form of a number of beeps. This was found to be inadequate in providing sufficient feedback to the user concerning errors made. Also the feedback provided was found to be inconsistent as different numbers of beeps were heard when the same error (such as attempting to delete waypoints using the incorrect button) was made in varying display modes (map display versus flight plan display).

Completion of certain tasks was found to have been "hampered or prevented" (p. 7) due to inconsistencies in GPS menu structure and button function allocation (Wreggit & Marsh II, 1996). Inconsistencies also were found to be a problem concerning the on-screen help option. This feature was not available for all the screens and, certain words such as "escape" and "exit" were used interchangeably for the same function in some of the help screens.

Finally, head-down glance time was found to be considerable, calculated as at least 10 seconds, on average, per participant (Wreggit & Marsh II, 1996). As described previously, this can be quite hazardous considering the fact that "time away from scanning the outside environment and the aircraft instruments reduces situational awareness" (Wreggit & Marsh II, 1996, p. 8) and this could potentially lead to mid-air or CFIT accidents.

The aforementioned design flaws are not just limited to the GPS unit used in the study by Wreggit and Marsh II. Indeed, a lack of due consideration to ergonomics and basic human factors principles has meant that many GPS devices currently being used in aircraft have not been optimized from a usability standpoint (Heron et al., 1997). Heron and colleagues provide a detailed discussion of the human performance concerns that must be considered when designing a GPS interface for pilots.

Heron et al. (1997) described many deficiencies in design that hamper effective interaction between the GPS receiver and the human operator. Among these deficiencies are those that induce keystrike problems whereby buttons are inadvertently pushed as a result of projecting key contours and/or insufficient spacing between buttons. This problem is usually exacerbated in-flight by the fact that pushing the wrong button can wreak havoc with the existing status of the display by, for instance, placing the user on a new screen with no relevance to the task. It is easy to see how the levels of confusion and frustration can mount, especially when workload levels are relatively high (during an instrument approach, for example) as there is usually no "undo" button to revert to the previous status of the display, requiring the pilot to perform many steps (button pushes) just to return to the previous screen (Heron et al., 1997).

Undue load on memory was another issue identified by Heron et al. (1997). This problem stems from the requirement for the pilot to memorize the many steps required to perform simple navigation tasks using the GPS device. For instance, Heron et al. (1997) found that the number of steps required to load a nonprecision instrument approach varied between 8 and greater than 30 steps depending on the model of unit used. The number of steps required has serious implications for head-down time, workload, and situation awareness. Memory load is further increased when the user must remember the many functions the individual knobs and buttons may have depending on screen displayed. Loss of memory during critical segments of flight can lead to excessive head-down time and confusion (Heron et al., 1997).

Counterintuitive logic is another problem that negatively impacts memory load (Heron et al., 1997). The nature of many GPS units used for aerial navigation is such that there is a limited number of buttons/knobs present on the hardware interface. Given this limitation, many devices require the use of the same buttons for performing different functions. However, problems arise when, as discussed previously, the labeling of some of the buttons is non-intuitive for

certain functions. An example is the requirement to press "select" to exit from a particular mode (Heron et al., 1997). Such labeling may actually impede memorization and increase workload as pilots attempt to retrieve the information from their long-term memory.

Concerning the actual depiction of characters on displays, Heron et al. (1997) identify three specific principles that appear to be neglected by many GPS manufacturers. These are: detectibility, recognizability, and readability. Detectibility refers to how well characters can be distinguished from their backgrounds. Recognizability refers to the ease with which characters can be discriminated. Finally, readability refers to how well character groups "convey meaningful information" (Heron et al., 1997, p.15). These three principles are especially important with regard to annunciation. Indeed, when important messages are provided by means of on-screen annunciator lights, they must be easily discernable so that situation awareness may be enhanced.

Another issue of concern is whether a graphical display form is superior to a textual display form. Graphical displays typically incorporate the use of shapes and colors to convey meaningful information to pilots while textual displays rely on human abilities to read and interpret verbal information presented. While both forms provide information visually, the most appropriate method of presentation should be based on the type of information presented. For instance, verbal displays may be more suitable for presenting abstract information while pictorial representations may be more suited to conveying complex spatial patterns (Wickens & Hollands, 2000). In some cases, however, when information processing times are critical, a redundant combination of graphical and textual information may be more useful and work better than either element employed singularly (Wickens & Stokes, 1988).

Human Operator Issues

Unsuccessful or problematic interactions with GPS interfaces may not be a result of just poor interface design, but also may be caused by the human operator himself/herself. For instance, an operator that has little or no familiarity or training with a particular device may find it difficult to perform a function (such as loading an approach) because of the lack in knowledge of or little previous exposure to the device (Adams et al., 2001). Additionally, the multitasking abilities of the operator (pilot) also may play a role. For instance, pilots with high levels of flight experience and/or who are trained to deal with high task load operations (e.g. airline transport pilots, flight instructors, etc.) may find it easier to adapt to or manage the additional challenge of, say, loading an instrument approach into a newly installed GPS unit, than less experienced pilots whose training may be limited to simple routine aerial maneuvering/navigation operations (e.g. student pilots, private pilots, commercial pilots, etc.).

In summary, it must be stated that the many human factors issues identified above concerning GPS unit interface design are by no means exhaustive. How-

ever, they do provide an indication of the many aspects of human performance that must be considered when designing such interfaces. Finally, the available research has thus far shown that, while GPS devices can be valuable tools for pilots, an awareness of usability issues and common pilot-induced errors will be essential both for the pilots themselves and the GPS manufacturers.

Statement of Problem

With the continued emphasis on GPS technology in aviation comes the responsibility of the aviation community in general to ensure that the devices used to capitalize on this technology, namely the GPS receiver hardware and software interfaces, are designed in a manner that truly enhances safe and efficient aircraft operation. It has been shown that manufacturers of GPS units currently in existence have, in some instances, failed to follow simple human factors principles to the detriment of the end user, the pilot. Relatively few studies have been conducted to date concerning GPS receiver interface design and usability.

The current lack of research in this particular area and the absence of industry standardization with regard to GPS interface design have created a void that can only be filled by investigative studies into safe, efficient, and effective interface design. This particular study aims to do just that. Specifically, this study was developed with the ultimate goal of providing recommendations pertaining to interface design for cockpit-mounted GPS receivers. Two different types of interfaces were compared. Additionally, the effect of certification level on pilot performance was investigated also.

Method

Participants

Thirty-two male participants participated in this study. However, one participant was unable to complete substantial portions of the experiment tasks and, therefore, had to be eliminated from the study (upon de-briefing it was discovered that he "wasn't feeling very well" i.e. the participant was ill). Participant ages ranged from 19 to 24 years. Approximately half (16) of the participants were pilots holding Certified Flight Instructor (CFI) certification while the remaining 15 were pilots with at least a Private Pilot certificate and an Instrument rating and/or pilots holding Commercial Pilot (IPC) certification but who had not attained Certified Flight Instructor certification. The pilots with Certified Flight Instructor certification were selected mainly from a pool of Florida Institute of Technology (FIT) Aviation LLC employees, while the pilots with at least a Private Pilot certificate and an Instrument rating were selected from a pool of FIT School of Aeronautics (SOA) students. The average total flight experience for CFI pilots was approximately 533 flight hours, and ranged from 250 flight hours to 1300 flight hours. The mean total flight experience of the IPC pilots was approximately 240 flight hours, and varied from 220 flight hours to 347 flight hours. All pilots possessed at least a third class FAA medical certificate.

With respect to GPS experience, only 9 (29%) of the 31 participants had ever specifically previously used either of the presented GPS interfaces. However, most of the participants, approximately 94%, reported that they did have prior experience with the Garmin 430 GPS receiver unit, a precursor to the Garmin 530 unit. None of the participants stated that they had received any formal GPS training at all.

As incentive to participate, a $150 cash prize was offered and was to be given to a winning participant on the basis of a random drawing conducted at the completion of the study.

Apparatus

The experiment was conducted at the FIT ARL (Applied Research Lab). Two IBM-compatible Portable Computers (PCs) at the ARL were used for the experiment. The monitor displays of each PC were placed side by side such that minimum gap existed between the two screens. One PC was designated to run NASA's MAT (Multi-Attribute Task Battery) software (Parasuraman et al., 1991) while the other PC ran ASA's GPS Trainer v. 2.0 © program (Aviation Supplies and Academics, Inc.) as well as NASA's TLX program (Hart & Staveland, 1988). The MAT software was selected for this study, as opposed to other flight simulation software currently available, because the MAT contained algorithms designed to produce specific outputs pertaining to user performance. These outputs provided for the accurate collection of analyzable metrics pertinent to this study.

The two GPS interfaces selected for this study were the Garmin 530 and the King KLN 89B. The selection was based on the fact that these particular units represent two completely different designs of cockpit-mounted GPS units (a graphically oriented interface on the Garmin 530 versus a textually oriented interface on the KLN 89B). Figure 1 is a depiction of the Garmin 530 interface and Figure 2 is a depiction of the King KLN 89B receiver interface. Both Figure 1 and Figure 2 are depictions to scale and represent one-half the size of the actual units. The GPS simulators were made available from ASA's GPS Trainer v. 2.0 © software.

Figure 1. Garmin 530 receiver interface

Figure 2. King KLN 89B receiver

The PC on which the MAT software ran was equipped with a Logitech Trackerball and a standard keyboard while the other PC was equipped only with a standard PS/2 mouse input device.

A post-experiment survey devised by the experimenter and containing Likert-scale and open-ended questions was used to obtain participant data concerning interface usability issues.

Procedure

Each experiment session was separated into three segments. The first segment constituted a *Tutorial and Practice* portion, and the second segment constituted performance of *Experiment Task 1*. The third segment constituted the performance of *Experiment Task 2*.

The *Tutorial and Practice* portion was designed to familiarize the participants with the tools to be used and the tasks to be performed. The first task assigned to each participant during this portion was familiarization with the MAT tasks to be performed during the experiment sessions. Specifically, this involved a five-minute session within which the participant simultaneously performed a monitoring task and a tracking task. These tasks were selected primarily because of their simulation of functions normally performed by pilots in-flight, namely, the scanning of instruments and gauges and the control of aircraft attitude and flight path. Thus, these tasks are referred to as flight-related tasks. Figure 3 is a depiction of the MAT display containing the monitoring and tracking windows.
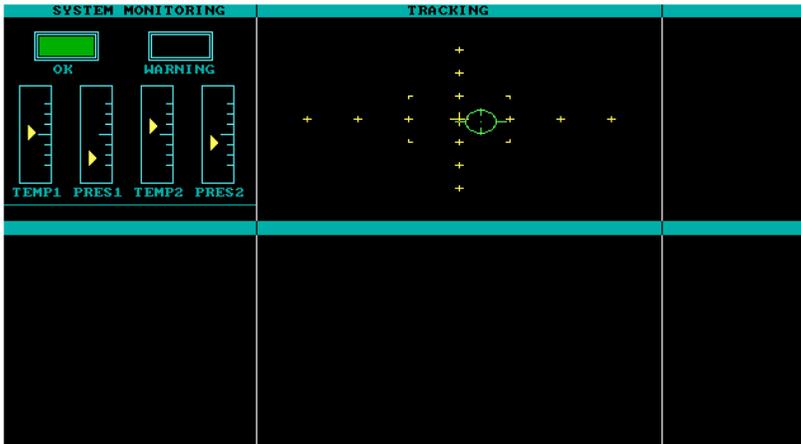
Figure 3. MAT display showing the two of six MAT panes used in this study.

The monitoring task required the participant to monitor four vertical gauges labeled Temp1, Pres1, Temp2, and Pres2 representing pressure (Pres1 and Pres2) and temperature (Temp1 and Temp2) variations of two aircraft engines. The pointers on each gauge would normally fluctuate within one scale deflection of the center marker. However, occasional random "system malfunctions" would occur and were indicated by the pointer on one of the four gauges shifting more than one scale deflection from the center position. For instance, the *System Monitoring* window depicted in Figure 3 illustrates a situation where the PRES1 gauge is "out of limits," while the other gauges remain "within limits." Participants were required to detect the malfunction and correct it by pressing one of four function keys (located on a keyboard) corresponding to the appropriate gauge. The faults would be corrected automatically 10 seconds from the onset if the participant failed to detect and correct the malfunction (a miss). A false alarm was recorded if a function key was depressed when no malfunction had occurred.

The performance measures pertaining to the monitoring task included detection rate of malfunctions or correct responses (CR), mean reaction time for detection and correction (MRS), and the number of false alarms (FA).

The tracking task involved the use of the trackerball device to control the motion of a green circular cursor representing the flight path of the aircraft. The cursor was programmed to move within a central boxed area in x and y directions according to a particular forcing function embedded within the MAT code. The participants were instructed to use the trackerball to attempt to maintain the cursor over a crosshair located in the center of the tracking display (see Figure 3).

The performance measure pertaining to the tracking task was a value derived from the x and y deviations of the cursor over time. This value represented the average root mean square (RMS) error score for the length of the session. This score will henceforth be referred to as a Mean Tracking Error (MTE) score.

The MAT software also produced an output of MAT run time, which measured the duration of the MAT session.

The difficulty level of the MAT was set to "normal," which required that the participant constantly monitor and correct for tracking deviations while also monitoring and correcting system malfunctions. This setting produced a situation not unlike piloting a single-engine aircraft under Instrument Meteorological Conditions (IMC) weather conditions (requires emphasis on instrument scan technique) in slight turbulence (requires deliberate attention to aircraft control).

Once the MAT tasks were completed to the required proficiency (i.e., a Mean Tracking Error RMS score of no greater than 290, and a detection rate of at least 66%), the participant completed a tutorial pertaining to each GPS task to be performed. The minimum scores required to demonstrate proficiency were de-

rived from an analysis of test subject data. The tutorial and practice portion of the session prepared participants to perform particular VFR and IFR scenarios, *Experiment Task 1* and *Experiment Task 2* respectively. The order in which the GPS interfaces were presented and used was counterbalanced between experiment portions (i.e. the tutorial, *Experiment Task 1*, and *Experiment Task 2*) and between participants to control for order effects. Additionally, separate t-tests comparing dependent measure results within each experience level category (IPC and CFI) were conducted to determine whether order effects did in fact exist. None of the t-tests were found to be significant.

Experiment Task 1 provided a typical VFR (Visual Flight Rules) flight-scenario in which the pilot was to perform certain GPS tasks while simultaneously performing the given flight-related tasks. The scenario presented was one in which the pilot had to initially proceed on a direct route to an airport of intended landing. Next, the pilot had to use his GPS receiver to find a suitable alternate airport with an appropriately lengthy runway given that the distance to the original destination was too great to allow for an immediate (or almost immediate) precautionary landing due to a possible system malfunction. Once the new airport was found, the pilot was asked to proceed directly to that airport with the help of the GPS unit by activating the "direct-to" function on the GPS (Note: the MAT tracking task functioned independently of the GPS routing changes). Finally, the last interaction with the GPS unit involved canceling the new direct-to routing given that the system problem had been resolved and a precautionary landing was no longer necessary.

Experiment Task 2, on the other hand, provided the pilot with a typical IFR (Instrument Flight Rules) scenario in which the pilot was instructed to proceed on a direct-to route to a particular airport and then load a given approach into the flight plan of the GPS receiver unit. While performing these instructions, the pilot, once again, had to simultaneously perform the flight-related tasks just as he/she would in an actual aircraft.

The tutorial ran directly from pertinent sections of ASA's GPS Trainer v. 2.0 © software. Specifically, the tutorial explained in detail, how each GPS receiver unit was to be used in the most appropriate manner in which to complete the given task.

After completion of the tutorial, the participants were provided with an opportunity to practice what had been learned. The practice portion of the experiment required that each GPS task be performed on each GPS simulator immediately after completion of the section of the tutorial pertaining to that particular task/GPS interface combination. The ability of the participant to accurately complete the given task was assessed and recorded and an *Accuracy and Completion* score was determined for the participant. This score was calculated based on the sum of assigned weightings provided for accurate entries and, if applicable,

a deduction penalty for failing to complete the given tasks within the allotted time. If the participant was unable to attain an *Accuracy and Completion* score of at least 70% within five minutes on each practice task (i.e., at least a score of 21 out of a possible 30 for Practice Task 1 and also at least a score of 14 out of a possible 20 for Practice Task 2), he/she was provided with an opportunity to review pertinent sections of the tutorial, and when satisfied, was allowed to try the practice portion again. Participants were allowed no more than two attempts to obtain sufficient proficiency. If a particular participant required more than 2 attempts, he/she was summarily dismissed from the study.

Once the *Tutorial and Practice* portion of the session was complete, each participant performed the *Experiment Session* portion. During the *Experiment Session*, the participants were required to perform GPS tasks (*Experiment Task 1* or *Experiment Task 2*) while simultaneously performing the flight-related tasks. Objective performance measures pertaining to the GPS tasks and flight-related tasks were recorded for each receiver interface. Subjective workload measures pertaining to each interface were obtained through the use of NASA's TLX program once participants had completed each separate portion of the *Experiment Session*. The NASA TLX measures included the following categories: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each measure was rated on a scale ranging from -1 for "low" or" good" (as appropriate) to 19 for "high" or "poor" (as appropriate). Upon completion of the entire *Experiment Session*, additional participant data were obtained through the use of a post-experiment questionnaire. The questionnaire contained certain Likert-scale questions to obtain subjective ratings of interface effectiveness for each experiment task. Such questions contained a scale ranging from "very poor" (with a value of 1) to "excellent" (with a value of 5). Additionally, the questionnaire was structured so that it also provided the thoughts of the pilots concerning each GPS interface, and how the interfaces related to the given tasks. Finally, the questionnaire also sought ideas on how the interfaces could be improved.

Experimental Design

For the purposes of this study, a two-factor mixed design Analysis of Variance (ANOVA) hypothesis test was conducted using *STATISTICA for Windows* (StatSoft, 1999). A separate A N O VA was performed for each dependent measure. The dependent measures included task completion times (time required to complete *Experiment Task 1* and *Experiment Task 2*), tracking error scores, and monitoring values. The independent measures consisted of: certification level (1) [Certified Flight Instructors (CFI) vs. Instrument-rated Private or Commercial Pilots (IPC)] and, (2) GPS receiver interface [Garmin 530 (G) vs. KLN 89B (K)]. Thus, a 2 x 2 mixed factorial design was conducted in which certification level was the *between-participants* variable and GPS interface used was the *within-participants* variable. An alpha level of .05 was used in all cases (á = .05).

Results and Discussion

Before we can discuss the findings of this study, we must first understand their context. Specifically, this requires a review of each of the experiment tasks and what they represented.

Recall that the tracking and system monitoring MAT (Multi-Attribute Task Battery) tasks represented typical flight-related tasks performed by pilots, namely, aircraft attitude and flight path control (represented by the tracking task), and flight instrument or system-gauge monitoring (represented by the system monitoring task).

For each separate scenario performed, the collected data (completion time, mean tracking error score, system monitoring values, etc.) pertained to the period spanning the initial immediate interaction with the GPS interface up until the completion of the final interaction with the interface (at which point the MAT session was simultaneously terminated).

The analyses of completion time results (recorded in minutes) for *Experiment Task 1* and *Experiment Task 2* indicated that there was no significant difference between CFI pilots (*Experiment Task 1*: M = 3.89, SD = 2.54; *Experiment Task 2*: M = 2.07, SD = 1.13) and IPC pilots (*Experiment Task 1*: M = 4.82, SD = 5.42; *Experiment Task 2*: M = 3.56, SD = 3.45) with regard to completion time (p > .05 for *Experiment Task 1 and Experiment Task 2*). This implies that, even though the CFI pilots had attained a higher level of pilot certification than IPC pilots, they were unable to complete both the VFR (*Experiment Task 1*) and IFR (*Experiment Task 2*) scenarios presented any faster than their lesser-rated counterparts. Therefore, it can be said that, with respect to speed of performance, level of certification did not provide one group of pilots with an advantage over the other group in this case.

With regard to the post-experimental questionnaire results for the VFR scenario, the Garmin 530 appeared to be the preferred interface overall, achieving a mean rating of 3.77 while the KLN 89B achieved a slightly lower mean rating of 3.45 (ratings based on a scale of 1 to 5 where 1 = poor, and 5 = excellent). The pilots, however, also commented that they preferred the instant display of pertinent information on the KLN 89B upon activation of the "direct-to" function, whereas additional steps were required to produce similar information on the Garmin 530.

With regard to the completion time results for the VFR scenario (*Experiment Task 1*), the main effect of GPS interface was also not significant (p > .05). This indicated that the pilots performed just as rapidly on the Garmin 530 interface (M = 4.12, SD = 2.89) as they did on the Bendix/King KLN 89B interface (M = 4.58, SD = 5.08). This was rather surprising considering the fact that the two interfaces were very different from one another, one being more "text-intensive"

(the Bendix/King KLN 89B), while the other one was more "graphics-intensive" (the Garmin 530). If completion time can be used to gauge the relative ease or difficulty of use of an interface, with shorter times indicating ease and longer times indicating difficulty, this result may be interpreted to mean that, in general, one interface may have been just as easy to use as the other; or alternately, the interfaces were equally difficult to use. The interaction between certification level and GPS interface used for the VFR scenario was not significant, p > .05. A further repeated-measures t-test comparison of GPS task completion times on the pre-experiment (practice) version of the VFR scenario also indicated that no significant differences in completion time existed between the two interfaces, p > .05 (two-tailed).

With regard to the post-experimental questionnaire results for the IFR scenario (*Experiment Task 2*), the Garmin 530 again appeared to be the preferred interface overall, achieving a mean rating of 4.00 while the KLN 89B achieved a relatively lower mean rating of 2.94 (ratings based on a scale of 1 to 5 where 1 = poor, and 5 = excellent).

The analysis of the IFR scenario completion time results did, however, reveal the main effect of GPS interface as significant ($F(1,29) = 7.95$, $p < .01$), with pilots completing the given scenario much more rapidly on the Garmin 530 interface (M = 1.62, SD = .83) than on the Bendix/King KLN 89B interface (M = 4.01, SD = 3.89). Additionally, the interaction effect between certification level and GPS interface was significant, $F(1, 29) = 5.49$, $p < .05$, (see Figure 4). The interaction indicated that CFI pilots actually required slightly more time than IPC pilots to complete *Experiment Task 2* on the Garmin 530 interface but were,
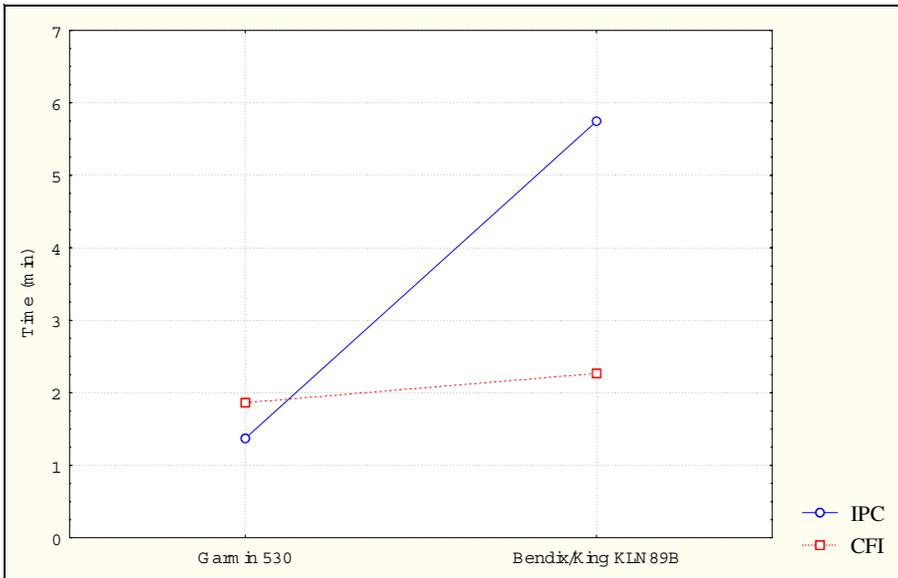
Figure 4. Interaction effect between certification level and GPS interface

however, able to perform the given scenario much more rapidly on the Bendix/King KLN 89B than the IPC pilots using the same interface (i.e., the Bendix/King KLN 89B). Further investigation of the interaction via a Bonferroni type correction (adjusted p = .0125) also revealed that the difference in IPC pilot performance between the two interfaces approached significance, $t(14) = +1.76$, $p = .025$, two-tailed.

This finding indicated that the lesser rated pilots (IPC pilots) had less difficulty with performance of the IFR scenario on the Garmin 530 than on the Bendix/King KLN 89B (Note: participants were instructed to complete the given tasks in the least amount of time possible, giving equal weight to tracking and GPS tasks; for almost all participants, this was observed to be the case). It thus follows that, for IPC pilots, head-down glance time (i.e., time spent tending to the GPS display at the expense of paying attention to the "aircraft" and its environment) must have been greater when the Bendix/King KLN 89B interface was used than when the Garmin 530 interface was used. The importance of this finding cannot be overlooked given the safety implications of long head-down glance times as described previously. Another possible implication of relatively long completion times is that this may reflect lengthy information processing times required by a particular interface. A repeated-measures t-test comparison of the interface completion times on the pre-experiment (practice) version of the IFR scenario indicated that GPS task completion times, and thus information processing times, for the textually-intense KLN 89B interface (M = 1.56 min, SD = .196 min) were significantly greater than those times for the more graphically oriented Garmin 530 interface (M = 1.05 min, SD = .183 min), $t(28) = -2.87$, $p < .01$, two-tailed.

Figure 4 also indicated that CFI completion times on each interface were fairly similar (the difference was found not to be significant, $p > .05$). This suggested that CFI pilots, perhaps as a result of their flight experience, did not have as much difficulty in adapting to the interface change as the IPC pilots.

An analysis of the Mean Tracking Error (MTE) score results for both the VFR and IFR scenarios revealed that the main effect of certification level was significant, (VFR scenario: $F(1,28) = 8.54$, $p < .01$; IFR scenario: $F(1,29) = 8.25$, $p < .01$). Specifically, CFI pilots (VFR scenario: M = 134.00, SD = 31.84; IFR scenario: M = 117.82, SD = 27.72) had significantly lower MTE scores than IPC pilots (VFR scenario: M = 182.00, SD = 64.42; IFR scenario: M = 161.87, SD = 58.90). Given the fact that the tracking task was likened to controlling the pitch attitude and flight path of an aircraft, it may be said that the CFI pilots were able to maintain better control of the "aircraft" (simulated) than IPC pilots while performing the required functions. This seemed hardly surprising given the fact that, unlike IPC pilots, CFI pilots, in earning their CFI certification, have been trained to allocate sufficient resources to flying the aircraft while simultaneously providing flight instruction to students. Thus, one might expect that the multitasking abilities of CFI pilots would be superior to those abilities of lesser rated pilots. Evidently, this was the case.

However, still very interesting, the VFR scenario MIE score results indicated that the Garmin 530 interface (M = 157.08, SD = 51.38) and the Bendix/King KLN 89B (M = 156.40, SD = 44.89) interface affected pilot "aircraft control" capabilities similarly (p > .05). This implied that, even though the interfaces were very different from one other, pilot abilities to perform the tracking task (or to "fly the airplane") remained the same for the VFR scenario regardless of GPS used. The interaction between pilot certification level and GPS interface was also found not to be significant for the VFR scenario, (p > .05).

With respect to the main effect of GPS interface, the IFR scenario proved to be more discriminating than the VFR scenario. The results indicated that the differences between the interfaces (with regard to the number, and type, of steps required for task completion) were more pronounced for the IFR scenario than for the VFR scenario. This was evident from the fact that only the IFR scenario produced a significant main effect of MIE scores concerning the interfaces (F (1,29) = 6.92, p < .05). The IFR scenario MIE score results showed that pilots performed significantly better on the tracking task when the Garmin 530 interface (M = 130.73, SD = 45.79) was used rather than the Bendix/King KLN 89B (M = 146.06. SD = 52.03), F (1,29) = 6.92, p < .05. In other words, pilots were able to maintain better "aircraft control" when using the Garmin 530 interface. Recall once again that the pilots also completed the IFR scenario significantly faster on the Garmin 530 as well. As described previously, the head-down time implications were that the pilots had lower head-down times when using the Garmin 530 as opposed to the Bendix/King KLN 89B. The interaction effect was found not to be significant, p > .05.

Due to the relatively low sampling rate for the system monitoring task, only the VFR scenario produced system monitoring results that could be analyzed. This is because the algorithm used by the MAT program to generate random system malfunctions, began to produce the system malfunctions only after some unspecified period of time had elapsed after the MAT session began. Thus, in several cases, participants completed the IFR scenario before the algorithm could generate a system malfunction. This did not occur with the VFR scenario where completion times were great enough that system malfunctions were generated for all participants. An analysis of the VFR scenario results revealed that neither the certification level main effect nor the GPS interface main effect was found to be significant. The indication provided by the result of the first main effect is that IPC pilot detection rates (M = 59.76, SD = 39.36) were similar to CFI pilot detection rates (M = 46.55, SD = 42.32), p > .05. This implied that CFI pilots and IPC pilots could not be distinguished based on system-gauge monitoring performance. This showed that while CFI pilots on average had greater total flight experience, they did not detect any more system malfunctions than the less experienced pilots for the VFR scenario. Since the main effect of GPS interface was not significant either (p > .05), the indication was that the pilots, on average, experienced no change in detection rate on each subsequent session whether completing the VFR scenario first on the Garmin 530 interface (M

= 52.35, SD = 41.96) and then the Bendix/King KLN 89B interface (M = 53.96, SD = 43.09) or vice versa. The interaction effect was not significant, p > .05.

The results of each of the 6 NASA TLX workload measures, for both the VFR and IFR scenarios, indicated no significant main effect of certification level differences in any of the subjective ratings between CFI pilots and IPC pilots. Similarly, for the VFR scenario, the main effect of GPS interface also was not significant. These results indicated that the subjective levels of workload experienced by both pilot types (CFI and IPC) on both GPS interfaces (Garmin 530 and Bendix/King KLN 89B) were not dissimilar. This was a rather interesting finding given the fact that the questionnaire results indicated that many pilots appeared to be partial toward either one unit or the other. Perhaps the old adage that says, "do not judge a book by its cover" holds true here.

The IFR scenario, however, did produce a main effect of GPS interface with regard to one workload measure in particular, "frustration." The pilots indicated that they were significantly more "frustrated" when working on the Bendix/King KLN 89B (M = 7.0, SD = 4.72) interface than when working on the Garmin 530 interface (M = 8.9, SD = 5.94), F(1, 29) = 4.81, p < .05. As shall be discussed subsequently, certain aspects of interface design may have played a role in producing such results.

The results of both the VFR and IFR scenarios showed that some participants did, indeed, have difficulty in interacting with the interfaces as evidenced by the relatively long completion times on at least one of the interfaces, if not both interfaces. As stated previously, certain interface design aspects of the units appeared to contribute to the difficulty experienced.

Concerning the VFR scenario presented, for those individual participants that appeared to have difficulty with the Garmin 530, the observed cause was a failure to quickly remember how to obtain information about the nearest airports. The performance of this function required the use of the large scroll knob located to the right hand side of the interface. However, this was not obvious or clear to all participants as evidenced by the fact that one pilot failed to complete the Experiment Task 1 session with the Garmin 530 interface due to this issue. Additionally, it is worth noting that many participants, in their answers to the research survey, made statements requesting that this knob be replaced by a "nearest" button on this particular interface. This suggested modification, while it appears trivial, actually may significantly enhance the usability of this interface.

For those participants that appeared to have difficulty with the Bendix/King KLN 89B interface on this same VFR scenario, the observed cause was a failure to remember to pull the right-hand-side inner knob out once at the initial nearest airport page so as to enable further scrolling through the subsequent nearest airports. Once again, this was the subject of frequent negative com-

ments from the pilots concerning the Bendix/King KLN 89B interface. A possible alternative to this design would be to incorporate a separate button to serve the same function as pulling out the knob. However, this might not solve the problem unless the button is appropriately labeled to make its function self-evident.

For the IFR scenario, the results showed that the use of the Garmin 530 interface for task completion did not appear to present much trouble to the pilots. This was most likely the result of the very simple and, apparently, the intuitive design, allowing easy access to the appropriate instrument approach page and its menus. The required steps to perform the given function were few and only required the pilot to press the "PROC" (i.e., "procedure") button and then make the appropriate instrument approach selections on the resulting menus. According to the research survey results, the steps were found to be very straightforward and did not cause much confusion or major obstacles to task completion. However, some of the survey answers illuminated the notion that the use of a scroll knob for data entry purposes on both the interfaces (the Garmin 530 and Bendix/King KLN 89B), was not very efficient, at least according to some pilots. These pilots provided suggestions of possible alternatives such as keyboards, touch-screens, or keypads, similar to those found in modern airliner FMCs (Flight Management Computers), cellular phones and PDAs (Personal Digital Assistants). Indeed, any one of these types of devices could provide an astute interface designer with a suitable model from which to adapt a new input device suitable for a cockpit-mounted GPS system. However, to be a truly effective design, due consideration must also be provided to the nature of the environment in which the system shall be placed. This requires consideration of such factors as limited space available in a cockpit area or panel, key spacing given the possibility of turbulence-induced keystrike errors, key illumination for night-time use, and so on. Another possible alternative input method could be a voice-controlled system incorporated within the GPS interface itself to allow "hands-free" operation of the unit. This idea was suggested and deemed highly favorable by several pilots. In fact, one of the pilots actually stated, "..a voice-activated system would be an absolute godsend."

The use of the Bendix/King KLN 89B interface to complete the IFR scenario, presented some of the pilots with a separate problem (or perceived problem) unique to this interface. These pilots stated that there was just too much scrolling involved in finding the appropriate page required to load the given instrument approach. They also stated that the menu structure of the Bendix/King KLN 89B unit was not very intuitive and needed to be modified to make it easier to find the approach page, which, on the existing unit, was located on a screen labeled "ACT 8" with A C T being an abbreviation for the word "active." Given the evident confusion, a possible improvement to the existing design might be to incorporate a separate "approach" button to call up an instrument approach menu. This was, in fact, a common suggestion concerning this unit. However, it also must be stated that, during the study, there was simply not enough time to

thoroughly familiarize the pilots with all aspects of each unit. It is possible that, given a great deal more time and practice towards learning the units, the pilots would have performed better. That said, should it really be necessary to be thoroughly familiar with the interface of a particular GPS unit to enable one to perform a common function like loading an instrument approach? Many pilots who must constantly switch between aircraft (and thus GPS units) in say, a flight training environment, may not believe so.

With respect to improvements to existing GPS units in general, most of the pilots indicated, on the research survey, that they would like to see an "undo" button on a GPS interface. Such a button would "undo" the last function performed, operating similarly to the "undo" function on many word processing applications found today. The selection of such a button is not surprising given the fact that many pilots are already familiar with and appreciate the use of such a function in their personal computers and would thus like to be afforded the same convenience on a GPS interface. This would, indeed, be very helpful by eliminating the need for the pilot to perform many steps (button pushes) just to return to the previous screen (Heron et al., 1997). The availability of such a feature on the GPS interfaces used in this study may have produced different results.

The relatively low malfunction detection rates produced by the IPC (59.76%) and CFI (47.65%) pilots during the VFR scenario could quite possibly have been the result of the relatively high workload levels experienced by the pilots. While the exact cause could not be determined, these low detection rates are indicative of poor situation awareness and also could have been a function of the deficiency in scan patterns adopted by the pilots (especially the CFI pilots, who detected less than half of the malfunctions).

Finally, although none of the pilots had any previous formal GPS training, some of the pilots actually stated that they felt that formal coursework on the subject would have improved their performance. This illustrated the (perceived) importance of such training to pilots. However, the actual benefits gained may depend on the nature of the curriculum in terms of what is actually covered and how comprehensive the training is.

Conclusion and Recommendations

We have found that, for the given VFR scenario, both of the GPS interfaces (the Garmin 530 and the Bendix/King KLN 89B) used in this study affected pilot performance similarly with respect to task completion times and mean tracking error scores produced. Thus neither interface was superior based on pilot performance on the VFR scenario. However, for the IFR scenario, one interface was clearly superior from a usability standpoint, at least in some respects. This can be deduced from the fact that, with respect to this scenario, pilots generally performed better when using the Garmin 530 interface than they did when using

the Bendix/King KLN 89B. They completed the given tasks more rapidly on the Garmin 530 interface, and additionally, performed even better on the tracking task while using this interface. Also, for the IFR scenario, the Garmin 530 interface was found to be the less "frustrating" interface.

In conclusion, the results of this study indicated that, while the Garmin 530 interface was the superior interface for the IFR scenario, it was actually no better than the Bendix/King KLN 89B for the VFR scenario.

Recommendations

With the increased reliance of the aviation industry on GPS technology comes the need for establishment of appropriate guidelines pertaining to interface design. Such guidelines will be important in designing interfaces that will provide pilots with safe and effective tools for aerial navigation.

This section contains pertinent recommendations for cockpit-mounted GPS interface design based on the literature review, the research survey data, and the study findings. It is expected that the FAA, GPS manufacturers, and designers will be able to adapt these recommendations when developing the necessary guidelines.

Recommendation 1

The first recommendation pertains to the screen of the GPS unit. The screen is a very important aspect of a GPS interface. It currently represents the major method, and sometimes the only method, of presenting useful data to the pilot. Thus, it is recommended that the screen be large enough that necessary information can be displayed in a clear uncluttered manner, yet it should not be so large as to require a great deal of panel space. Survey respondents indicated that they appreciated the size of the screen on the Garmin 530. Thus, the size of the Garmin 530 screen may serve as a suitable example for future units.

Recommendation 2

The second recommendation pertains to the information presented on the screen. Too much data provided can overwhelm the pilot and yet too little can instigate endless searches for desperately needed information. For instance, when a pilot initiates a direct-to function to a particular airport, he/she may prefer to immediately be presented with the distance to the airport, the GPS course to the airport, the estimated time enroute, and so on. Other pilots may have their own preferences concerning the type and amount of information that should be displayed. So, to allow for varying pilot needs, a partially programmable interface (with a reasonable set of default settings) would be useful to allow pilots to select the kind of information that should be instantly available on the screen once a particular function has been activated. An adaptive interface also may be useful, with the interface designed to provide appropriate important information such as, the mode of flight (enroute, cruise, approach, etc.). Also important is how the information is presented; a lack of color, use of inappropri-

ate colors, and use of unsuitable fonts and font sizes can contribute to difficulty in reading the display, requiring the pilot to strain to obtain needed data. Due consideration also must be given to the menu structure of the unit. Required menus should be readily available and navigation to those menus should be intuitive. Additionally, any annunciators should be appropriately positioned on the display and distinguishable from the background so as to "grab" the attention of the pilot. Perhaps auditory annunciators also could be used to take advantage of the auditory channel available. These improvements could greatly enhance existing units. Thus, it is recommended that these factors be adequately considered when determining information presentation format.

Recommendation 3

The third recommendation pertains to the input method used for data entry into the GPS unit. Given the finding indicating that many pilots felt that the use of a scroll knob for data entry is not very efficient, it is recommended that alternative input methods be sought so as to decrease the reliance on the scroll knob. Some possible alternatives suggested previously are: use of keyboards, touch-screens, or keypads (with keyboards or keypad designs allowing for a fold-in capability so as to allow for maximum display space). However, due consideration also must be provided to the nature of the environment in which the system shall be placed. This requires consideration of such factors as limited space available in a cockpit area or panel, key spacing given the possibility of turbulence-induced keystrike errors, key illumination for nighttime use, and so on. Another possible alternative input method could be a voice-controlled system incorporated within the GPS interface itself to allow "hands-free" operation of the unit. Currently, this technology has been applied successfully within the automobile industry where high-technology GPS devices containing such features (e.g. the Pioneer AVIC-650 VT GPS) may be found in some luxury vehicles. However, adaptation of this technology for the aviation industry, while quite possible, presents a few challenges. Factors to be considered are: noise levels prevalent in typical single-engine and twin-engine propeller-driven aircraft, the amount of training required for the unit to accurately recognize user commands, how to activate the system to accept voice commands, and so on. A possible means of activating such a system only when necessary might be incorporating a separate microphone button onto the aircraft yoke but designated for communication to the GPS only.

Recommendation 4

The fourth recommendation pertains to the buttons available on the interface. The number of buttons should be neither excessive, such as the spacing of the buttons induces keystrike errors, nor so few buttons with multiple functions that undue load is placed on the memory of the operator. If multiple functions are necessary for certain buttons, the labeling should be clear and intuitive. Also, certain buttons such as "nearest," "approach," and "undo" could possibly be adopted as standard buttons on every cockpit-mounted GPS interface. Standardization would allow pilots to perform certain typically used function on a

GPS unit even though they may not be intimately familiar with the unit. A "nearest" function button would enable the pilot to instantly find the nearest airports, fixes, and navaids. An "approach" button would allow a pilot to instantly call up an approach, and an "undo" button would allow a pilot to "undo" the last function performed, helping the pilot recover from inadvertent mistakes. The "nearest" and "approach" buttons are ones that are commonly used by pilots and the "undo" button is expected to be highly useful according to this study. Therefore, it is recommended that these factors be adequately considered when determining how many buttons should appear on an interface and which buttons ought to be designated as standard buttons on every interface.

Recommendation 5

The final recommendation concerns the need for pilot training with respect to GPS use. While the results of this study indicated that the more experienced pilots were able to adapt more easily to a switch between two vastly different interfaces, this should not be interpreted to mean that flight experience level alone determines how well one will be able to interact with an interface. In fact, since many of the GPS units found currently are highly complex, a great deal of formal training may be required before any satisfactory interactions may occur. Also, even though GPS manufacturers can attempt to build safety into their units by reducing head-down time through intuitive design, pilots must still be educated about their own responsibilities concerning safe GPS use. This involves requiring pilots to undergo an FAA-mandated training program wherein they learn, among other requirements: (1) the limitations of GPS units (e.g., potential database errors, satellite reception problems, etc.), (2) how to properly divide attention between the aircraft and the GPS interface, and (3) the need to be thoroughly familiar with the GPS unit no matter how simplistic the interface appears to be. Additionally, pilots should be able to demonstrate proficiency at performing typical GPS interaction scenarios such as those provided in this study, while simultaneously flying the aircraft and paying attention to its environment. It is expected that these measures will encourage pilots to seek GPS training pertaining to their own specific units with qualified instructors. This shall enhance the level of safety and awareness of those GPS users who take to the skies.

Limitations and Areas for Future Research

This study represented an attempt to fill the present void in the area of GPS interface design for the aviation environment. While a great deal has been learned through this effort thus far, many avenues for future exploration have been identified. For instance, this study compared only two popular aviation GPS interfaces. However, there are many interfaces currently in existence that have not been compared on the basis of practicality. Future studies of similar design could compare multiple GPS units (perhaps six or more), and, additionally, the research setting could be slightly modified such that real GPS interfaces are used rather than simulated interfaces. Furthermore, the scope of this study

could be expanded in the future to include cognitive and behavioral task analyses associated with the use of different GPS units.

It should be noted that, in general, the use of personal computers to simulate actual aircraft components and scenarios (as was done in this study), while practical, limits the realism factor experienced by participants. There also is the additional issue of introduction of a potential confound to the study (participant computer experience levels). While this was not expected to be a major factor influencing the results, it is nonetheless possible that it could have played a role, however small. This should be taken into consideration in any future studies using similar apparatus. Further suggested is, for enhanced realism, an actual aircraft simulator be used in any follow-up studies.

Another area worth exploring is the development and evaluation of the effectiveness of one or more GPS training program(s), such as a study to compare a control group not receiving any formal GPS training with one or more group(s) receiving formal GPS training. The most effective program could then be submitted to the FAA for approval for use in flight schools.

Finally, a more extensive study, using a survey of thousands of pilots, could be designed to obtain pilot opinions on GPS interface design. The results could be tabulated and used to develop a GPS interface with the most commonly requested features. Then, in a study similar to the current one, the prototype GPS could be compared with existing units to determine what features do significantly affect pilot performance. It is expected that the results could further lead the aviation industry closer to the elusive "ideal" GPS interface that provides pilots with the safest and most effective tool for aerial navigation.

References

Adams, C.A., Adams, R.J., & Hwoschinsky, P.V. (2001). Analysis of adverse events in identifying GPS human factors issues. Volpe NTSC.

Aviation Supplies & Academics, Inc. *ASA GPS Trainer V.2.0* [Computer program]. Newcastle, WA.

Goel, S. (n.d.). Opportunities in aviation and GPS. Retrieved March 23, 2003, from: http://www.gisdevelopment.net/technology/gps/techgp0037pf.htm

Hart, S. G. & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam. The Netherlands: Elsevier.

Heron, R. M., Krolak, W., & Coyle, S. (1997). *A human factors approach to use of GPS receivers*. Heron Ergonomics, Inc. Vancouver, Canada.

Nendick, M. & St. George, R. (1995). Human factors aspects of global positioning systems (GPS) equipment: A study with New Zealand pilots. In *Proceed-*

ings of the Eighth International Symposium on Aviation Psychology (pp. 152-7). Columbus, OH: The Ohio State University.

Parasuraman, R., Bahri, T., & Molloy R. (1991). *Adaptive automation and human performance: I. Multi-task performance characteristics* (Tech. Report CSL-N91-1). Washington, DC: Catholic University of America, Cognitive Science Laboratory.

Parasuraman, R., Molloy, R., & Mouloua, M. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*, (4), 665-679.

StatSoft, Inc. (1999). *STATISTICA for Windows* [Computer program manual]. Tulsa, OK.

Wiener, E.L. (1988). Cockpit automation. *Human factors in aviation.* San Diego, CA: Academic Press.

Wickens, C. D. & Hollands, J. G. (2000). Attention, timesharing, and workload. *Engineering psychology and human performance* (3rd ed.) (pp.439-480). New Jersey: Prentice Hall.

Wickens, C. D. & Stokes A. F. (1988). Aviation Displays. In E. A. Wiener & D. C. Nagel (Eds.), *Human factors in aviation.* (pp. 387-421) San Diego, CA: Academic Press

Williams, K. W, Yost, A., Holland, J. & Tyler, R. R. (2002). *General aviation pilot use of ADS-B displays in bethel, Alaska: safety issues*. Retrieved March 24, 2003, from: http://www.akavcc.org/iaatc2002/papers/12%20GENERAL%20AVIATION%20PILOT%20USE%20OF%20ADS-B%20DISPLAYS%20IN%20BETHEL,%20.pdf

Wreggit, S. S. & Marsh II, D. K. (1996). *Cockpit integration of GPS: initial assessment-menu formats and procedures*. Human Factors Research Laboratory. FAA – Civil Aeromedical Institute. Oklahoma City, OK.

Development of a Small Aircraft Maintenance Monitoring System (SAMMS) Interface for In-flight Mechanical Fault Management

Andrine Y. Clarke, John E. Deaton, Nathaniel E. Villaire,

and

Muzaffar A. Shaikh
Florida Institute of Technology
150 West University Boulevard
Melbourne, FL 32901-6975

Abstract
In this study, two prototype interfaces were developed and tested for a system called Small Aircraft Maintenance Monitoring System (SAMMS). The two prototype interfaces, which provided more direct and detailed information about a failure, were compared to the current mechanical fault detection system in the Piper Warrior. The results of this study indicated which one of the prototypes was the more efficient way to display information generated by SAMMS. Moreover, results demonstrated that SAMMS improves performance by reducing diagnosis time of a mechanical failure and increasing the accuracy of that diagnosis (though not significantly).

Introduction

According to the National Transportation Safety Board online database, mechanical failure is a contributing factor to numerous general aviation (GA) accidents each year. Mechanical failures are indicated by a fault detection system that consists of three annunciator lights ("OIL," "VAC," and "ALT"), fuel gauges, oil gauges, and an ammeter. The fault detection system that notifies the GA pilot of in-flight mechanical problems is very limited. After conducting informal interviews with several flight students and instructors for background information, the following concerns were mentioned repeatedly. The primary alert fea-

ture of the system is the set of three annunciator lights. The lights are small and poorly illuminated, and the pilot is not always aware that there is a failure. There are instruments (i.e., the oil pressure gage on the Piper Warrior) that display incorrect readings. When a failure occurs, the pilot must diagnose the situation based on readings from gauges that are known to be faulty and from other sources such as sounds or vibrations. Using this system for fault diagnosis allows for errors that cause accidents.

A system that provides more direct and detailed information regarding the current state of vital systems or components could reduce errors, such as misdiagnoses, and alert the pilot to any potential malfunction. The Health and Usage Monitoring System (HUMS) is such a system. Stewart Hughes Ltd. in England and Teledyne Controls in the United States (U.S.) designed HUMS in response to concerns about helicopter operations over the North Sea. Since its implementation in 1991, HUMS has improved the safety of helicopter operations. One lethal accident was prevented when HUMS detected a cracked 10-mm bolt between the engine and gearbox in a Norwegian Commercial Chinook (Marsh, 1996).

HUMS provides detection, diagnosis, and prognosis information about failures through the use of sensors (i.e., accelerometers and chip detectors) located throughout the airframe and engine and advanced processors located in the cockpit interface that generates an output to the pilot. The Condition-based Maintenance (CBM) philosophy was the theoretical concept behind the HUMS. The objective of the CBM philosophy is to accurately detect the current state of mechanical systems and accurately predict systems' remaining useful lives (Deaton, Glenn, & Popp, 1999). Maintenance is performed based on the current condition of a part or system instead of on an elapsed period of time. Operators are able to perform maintenance as needed to prevent operational deficiencies or failures essentially eliminating costly periodic maintenance and greatly reducing the likelihood of machinery failures.

The U.S. Navy has been making the transition to the CBM philosophy by incorporating HUMS into selected helicopters after discovering that many of the rotor gears being replaced on helicopters were still good. As part of its research program, the Navy has developed and tested an interface called the HUMS Interface System (HINTS). HINTS is an electronic knee-board device connected to the Warning Caution and Advisory cockpit alerting display and the HUMS (Glenn, Deaton, Barba, & Popp, 2000). One Navy study assessed the value that various kinds of information have on the ability of the aircrew to manage in-flight mechanical faults, demonstrated the HINTS interface concept, and explored its potential benefits. The conclusions of the study indicated that: (a) HINTS improves fault diagnosis, (b) the analysis information provided by the HINTS was the most useful portion, (c) HINTS reduced workload for some pilots, and (d) had a neutral or beneficial effect on crew communication. The overall tone of the debriefings was very positive with respect to aircrew opinions on the usefulness of the HINTS (Deaton, Glenn, Popp, Barba, & Bowers, 1998).

In this study, the CBM philosophy was applied to GA by developing a hypothetical HUMS-like system referred to as the Small Aircraft Maintenance Monitoring System (SAMMS). It has the same operating principles as HUMS and offers improved fault management over the current fault detection system found in GA. The idea of applying a system like SAMMS to GA was suggested by Ritchie (1998) who recommended redesigning aircraft equipment to incorporate a computer that receives all the inputs that relate to an in-flight mechanical problem. The introduction of HUMS to helicopter operations in the North Sea has shown an improvement in performance and safety. Therefore, the introduction of SAMMS to GA is expected to have the same effect.

This study was an extension of research performed by the Navy, but instead of a helicopter, the Piper Warrior was chosen to represent the typical GA aircraft as defined by Turnbill (1999) in a National Aeronautics and Space Administration (NASA) report. The study had two main purposes. First, it determined an effective way to display information generated by the SAMMS to the GA pilot. This was achieved by comparing two prototypes of the SAMMS interface (SAMMS1 and SAMMS2) to each other and to the baseline (the current mechanical fault detection system in the Piper Warrior). Second, it determined if the information provided by the SAMMS improved performance by reducing the time it takes to diagnose an in-flight mechanical failure and increasing the accuracy of that diagnosis.

Method

There were two main parts to this study. The first part involved the use of a questionnaire that inquired about the pilot's experience with in-flight mechanical failures in the Piper Warrior. The questionnaire also provided information that introduced SAMMS and its capabilities. The second part involved the use of a computer program that compared three displays: the current fault detection system in the Piper Warrior as the baseline and two prototypes interfaces (SAMMS1 and SAMMS2).

Part One – The Questionnaire

Participants

Sixty-two participants (24 flight instructors and 38 flight students) filled out a three-page questionnaire referred to as Questionnaire #1. Of the 62 participants, five attended a private flight instruction school, and the other 57 participants were Florida Tech flight students. All participants were required to hold at least a private pilot's license and have flight experience in the Piper Warrior. The mean total flight times for the flight instructors and the flight students were 545 hours and 208.5 hours, respectively. The mean total flight times on the Piper Warrior for the flight instructors and the flight students were 347 hours ($SD = 323.57$) and 109.7 hours ($SD = 71.44$), respectively.

Questionnaire #1

Questionnaire #1 was divided into Parts I and II. Some of the questions had

a rating scale of "Very Poor" to "Excellent" (1 to 5). There were 11 questions in Part I regarding the mechanical failures experienced by the participants on the Piper Warrior. It also solicited the kind of improvements they would recommend. The following questions and rating scale are examples from Part I of Questionnaire #1.

(a) What are the most common mechanical failures you have encountered in flight on the Piper Warrior?

(b) How would you rate the current system used to diagnose a mechanical failure in-flight?
Very Poor _____ Poor _____ Fair _____ Good _____ Excellent _____

There were five questions in Part II that introduced participants to SAMMS, asked how much information the system should provide and whether the participants would find the additional information useful. Data from Questionnaire #1 was used in the development of the display comparison experiment.

Procedure
The questionnaire was distributed in a classroom setting at Florida Tech and the study and its objectives were briefly explained to the participants. Those that volunteered to participate in the second portion were instructed to write their names and telephone numbers in the appropriate space on the consent form. After the participants understood what was expected, they were instructed to sign the consent form and to begin completing the questionnaire. The completed questionnaire was then collected and participants who agreed to volunteer for the second portion were told that they would be contacted to schedule a convenient time for them to complete the display comparison experiment.

Part Two – Display Comparison Experiment

Participants
Twenty-five participants completed the display comparison experiment. Participants consisted of 14 flight students and 11 instructors from Florida Tech. All participants were required to hold a private pilot's license at a minimum and have experience flying the Piper Warrior. They also had a broad range of flight hour experience. The mean total flight times for the flight instructors and flight students were 755.5 hours and 130.91 hours, respectively. The mean total flight times on the Piper Warrior for flight instructors and flight students were 501.79 hours (SD = 445.82) and 94.55 hours (SD = 64.13), respectively.

Apparatus
The display comparison experiment involved the use of the following:
- a computer program that presented the three displays,
- a demo version of that program,
- NASA Task Load Index program (Hart, S. G. & Staveland, L. E., 1988) as a measure of workload,

- and another paper questionnaire for participants' opinions of the displays.

The computer programs were presented to the participants on a Gateway E-3100 computer with an Intel Pentium II processor and a 17-inch Gateway monitor.

The Display Comparison Program

The Visual Basic 6.0 compiler kit by Wang (1998) was used to create the computer program that presented the three displays. The program incorporated the three displays (the baseline, SAMMS1, and SAMMS2) in nine fault scenarios. A "fault scenario" refers to a situation involving a mechanical failure or malfunction. There were nine fault scenarios, three for each display. Incorporated among the fault scenarios were false alarms, e.g., the fuel pressure gauge reading zero when both fuel tanks are full. The addition of false alarms into the program increased its realism. The program allowed the order of the displays and the order of the nine fault scenarios to be randomized. This helped to reduce training or order effects.
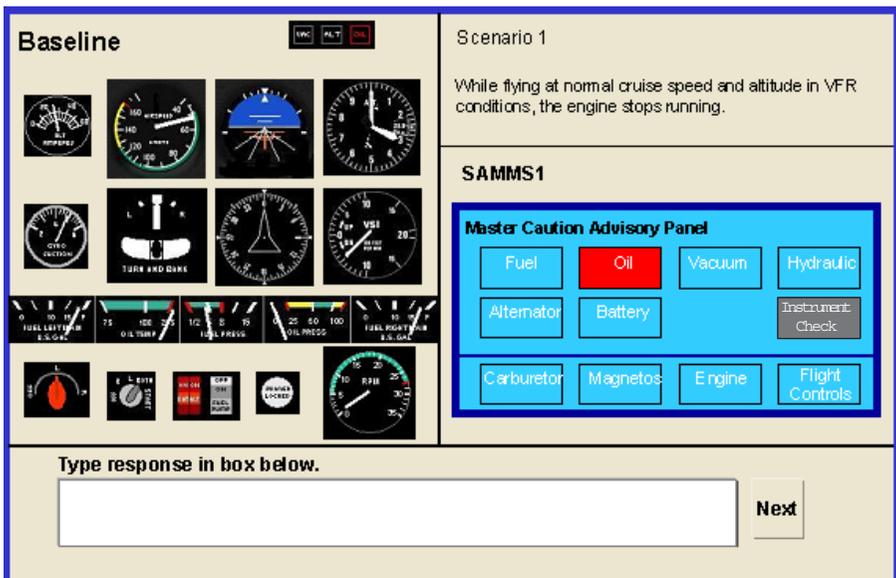
Figure 1. Layout of the display comparison program.

Figure 1 shows the basic layout of the program, which is split into four sections. The baseline (the current system displayed in a Piper Warrior) is located in the top left section and was present throughout the entire program since SAMMS supplements the current fault detection system rather than replace it. SAMMS offers the same information as the baseline, but it provides it more directly by using text instead of instrument readings. SAMMS also provides additional information, interprets instrument readings, and generates a brief sum-

mary to the pilot (see Appendix). SAMMS1 (see Figure 2) is in the middle on the right. SAMMS2 (see Figure 3) is located in the same position as SAMMS1 when it appears in the program. This display configuration allowed for comparisons between SAMMS1, SAMMS2, and the baseline. A brief description was provided for each fault scenario in the top right section (see Figure 1). The program did not have important sound clues, such as engine roughness or stoppage. When clues were relevant, they were presented in this section. After the participants determined the cause of the fault, they typed their responses in the bottom section. The "Next" button in this section allowed the participants to move to the next scenario.
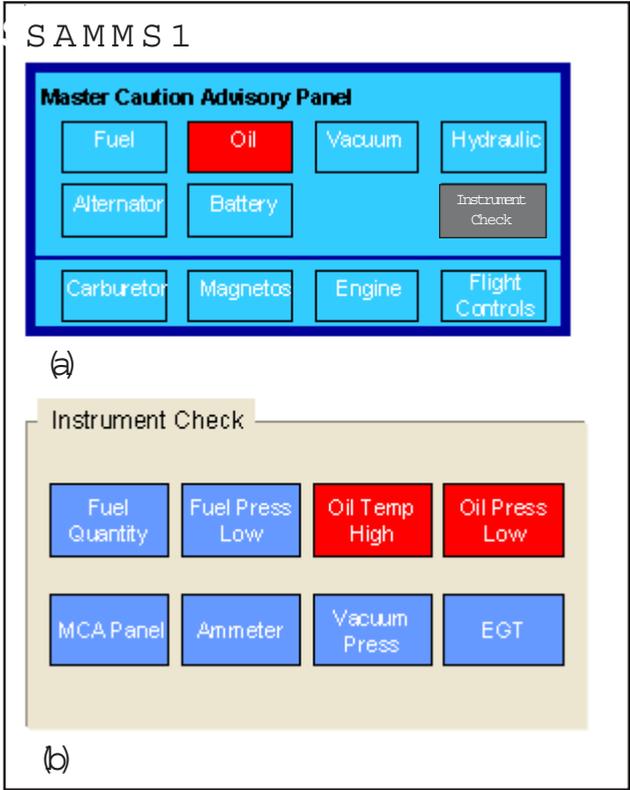
Figure 2. SAMMS1 – (a) MCA panel with instrument check button and (b) instrument check menu.

Displays. The baseline was a representation of the cockpit instrumentation system in the Piper Warrior that included the fault detection system of engine instruments and annunciator lights used for fault management (see upper left section of Figure 1). SAMMS 1 and 2 were prototypes of the SAMMS interface. SAMMS2 was based on the Navy HINTS prototype and consisted of a series of buttons with text. The first menu presented was the Master Cautionary Panel with buttons that turn red when there was a malfunction (Figure 2a). It offered
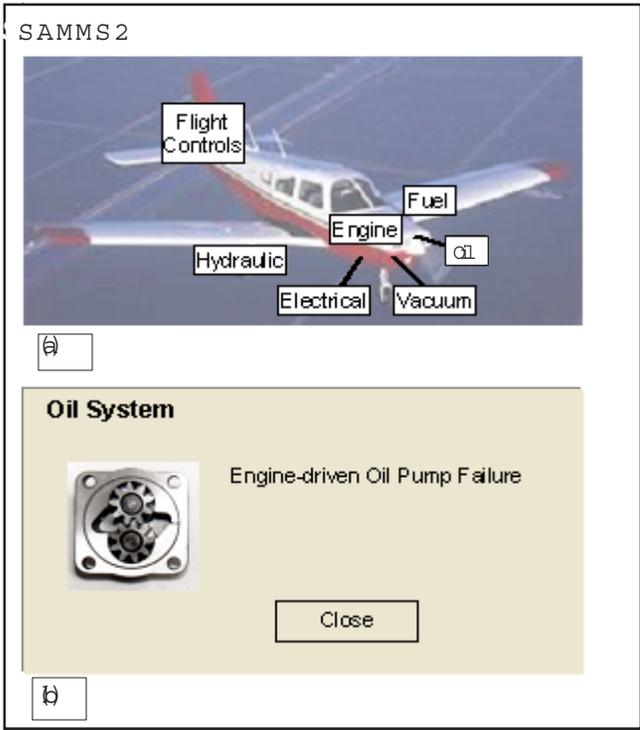
Figure 3. SAMMS2 – (a) Main screen with system buttons and (b) secondary screen with exact cause of failure.

the option of clicking the "Instrument Check" button to receive more information. The "Instrument Check" button brings up the Instrument Check menu, which allowed the pilot to click and determine the exact cause of a malfunction (Figure 2b). SAMMS2 has a background picture of the Piper Warrior with buttons that turned red to indicate a failure or a malfunction (Figure 3a). Once a button is clicked, the exact cause appears with a picture of the suspected component (Figure 3b). Each display cycled through a series of mechanical fault scenarios, and the participants were asked to respond as if they were in-flight and under specified conditions.

Demo. The demo was created using Microsoft Visual Basic 6.0. It was a simplified version of the program that allowed the participant to see exactly what the program looked like during the experiment. It also allowed for the demonstration of all aspects of the program, and featured two fault scenarios, one of which was a false alarm.

NASA Task Load Index. The NASA Task Load Index (TLX) is a rating procedure that provides subjective workload scores based on an average of ratings on six different subscales. The NASA TLX version 1.0 appeared as a simple program in which the participants entered their subject ID and then used the mouse

to click along a rating scale to indicate their workload level for each subscale. For the purpose of this study, only four of the six subscales were used: mental demand, effort, frustration level, and performance. The physical and temporal subscales were excluded because the experiment did not require any physical excursion or calculation on the part of the participant. The mental demand, effort, and frustration subscales have a range of –1 to 19 (low to high). The performance subscale ranged from good to poor (–1 to 19) but was recoded to so that good = 19 and poor = –1. The scores from the four categories were also averaged to give each participant an overall workload score.

<center>Questionnaire #2</center>

Questionnaire #2 was a paper questionnaire consisting of nine questions used to obtain subjective ratings of the three displays (baseline, SAMMS1, and SAMMS2) after the participants completed the program. The participants were asked to rate each display in three categories: (1) understandability, (2) usability, and (3) an overall rating. The rating scale was from "very poor" (with a value of 1), to "excellent" (with a value of 5). The ratings for the three categories were averaged to provide a score for each participant for each display. The results helped to determine which of the two prototypes (SAMMS1 and SAMMS2) the participants thought best displayed the additional fault information.

Procedure

The experiment began by having the participants sit in front of a computer on which the program was installed. Pictures of the three displays were next to the computer and the definitions of fault scenario, mechanical failure, and false alarm were posted on the wall above the monitor. The pictures of the displays aided in describing what the study would involve and served as a reminder of how each display looked when the subjects completed the NASA TLX. All the participants were given the same set of instructions. Each display was explained, with the aid of the demo, allowing the participant to use the mouse to click on the program and go through the menus of SAMMS1 and SAMMS2. To assess that participants fully understood how to use the program, they were asked a series of questions by the researcher. After the participants fully understood how to use the program, they were asked to begin the program.

The program presented the participants with nine fault scenarios, three for each display. They were instructed to look at the displays and determine the cause of failure in each scenario. After each scenario, they entered their response in the box provided at the bottom of the screen. Once the participants completed the program, they were shown how to use the NASA TLX. Each participant cycled through the NASA TLX three times, once for each display. The participants then were asked to complete Questionnaire #2. After completing Questionnaire #2, they were asked if they had any questions. The researcher was always present during the experiment.

Data Collection

The program included a timer that recorded the amount of time needed to assess the cause of a fault scenario. This was referred to as "diagnosis time" which began when each scenario was displayed and ended when the participant clicked the "Next" button to move to the next scenario. The time used by participants to type their responses was included since many participants also seemed to use this time to assess the cause of the failure. The program was also designed to save the responses typed by the participants in a database file. The participants' responses to the scenarios were reviewed to determine accuracy and diagnosis time for each display.

Statistical Design

Independent variables consisted of the displays (baseline, SAMMS1, and SAMMS2) and experience (flight instructors and flight students). The dependent variables assessed included diagnosis time and accuracy with a count of "1" for correct and "0" for incorrect assigned based on the text input. The variables were organized in a 3 x 2 mixed model ANOVA design matrix with displays as the within subjects variable and experience as the between subjects variable. Two ANOVAs were conducted, one for each dependent variable (diagnosis time and accuracy). All numerical data were analyzed using SPSS 10.1 (SPSS Inc., 2002) for Windows.

The accuracy and diagnosis time data were stored in an Excel spreadsheet. For each display, there were three fault scenarios. Therefore, there were 18 responses per participant (9 accuracy and 9 diagnosis time responses). Each correct response was counted as one, for a maximum score of nine for each participant.

Results

Part One – The Questionnaire

Magneto failure was the most common mechanical failure experienced by pilots during flight. Magneto and alternator failures accounted for 49 percent of the common mechanical failures experienced on the Piper Warrior (see Appendix). The remaining seven failures, accounted for 51 percent of the common mechanical failures. Forty-two of 62 participants rated the current fault detection system in the Piper Warrior as fair. Most of the participants indicated they thought the system should be improved. The most common improvements suggested by participants were more annunciator lights and an increase in the reliability/accuracy of the engine instruments.

Twenty-three flight students and 13 flight instructors rated the current system as fair (see Appendix). The results also indicated how much information the participants thought the SAMMS should provide (see Appendix). Twenty-five flight students and 17 flight instructors selected the "min plus" option, which

means the SAMMS should provide minimum required information with the option of clicking for more details. SAMMS1, which was based on the Navy's HINTS, would be an example of an interface displaying a "min plus" amount of information. In addition, all the participants indicated that they thought the information provided by SAMMS would be very useful.

Part 2 – The Display Comparison

A 3 x 2 mixed model ANOVA (three displays and two experience levels) was used to analyze the diagnosis time and accuracy data (alpha level = 0.05). A Geisser-Greenhouse correction was used to correct for violations of the sphericity assumption.

A significant difference between experience levels in the diagnosis time data was found (F (1, 23) = 5.33, p = .03). The means and standard deviations for the diagnosis time data are indicated respectively in Table 1. Flight instructors diagnosed failures more rapidly than students did. This was expected since the instructors in this study had approximately three times the total flight hours (particularly flight hours on the Piper Warrior) as the flight students. Consequently, this extra experience may explain their abilities to diagnose failures more quickly.

Table 1.
Means and SDs for Diagnosis Time Data in Seconds

	Baseline	SDs	SAMMS1	SDs	SAMMS2	SDs	Cum. Means	Cum. SD
FI	36.60	(14.36)	29.12	(7.36)	34.69	(10.73)	33.47	(8.00)
FS	43.88	(15.58)	41.97	(13.41)	45.12	(17.63)	43.66	(13.87)
Cum. Means	39.80		34.77		39.28			
Cum. SD		(15.66)		(12.51)		(15.41)		

There was no significant difference in diagnosis time between the displays (F (2, 46) = 1.54, p = .23). However, this non-significance is very important. It shows, even with the addition of the new SAMMS interface to the diagnosis process, that it did not take a longer time to interpret the cause of the failures. The participants had only a brief experience with the two SAMMS prototypes, and yet did as well as with the baseline, which is certainly more familiar given that they would have used such a display when flying. With more time to practice, they may have done better with SAMMS.

There was no significant interaction effect for diagnosis time noted between the displays and experience levels (F < 1).

Analysis of the accuracy data indicated there was no significant difference between experience levels (F < 1). The means and standard deviations for the accuracy data are indicated respectively in Table 2. This was surprising since

the flight instructors were more experienced. Close observation of Table 2 showed that there was a relatively large difference between instructors and students for SAMMS1. However, this difference was not significant and should be interpreted cautiously.

Table 2.
Means and SDs for Accuracy Data

	Baseline	SDs	SAMMS1	SDs	SAMMS2	SDs	Cum. Means	Cum. SD
FI	1.71	(0.79)	2.93	(0.26)	2.64	(0.48)	2.43	(0.40)
FS	2.00	(0.60)	2.55	(0.66)	2.73	(0.45)	2.42	(0.54)
Cum. Means	1.84		2.76		2.68			
Cum. SD		(0.75)		(0.52)		(0.48)		
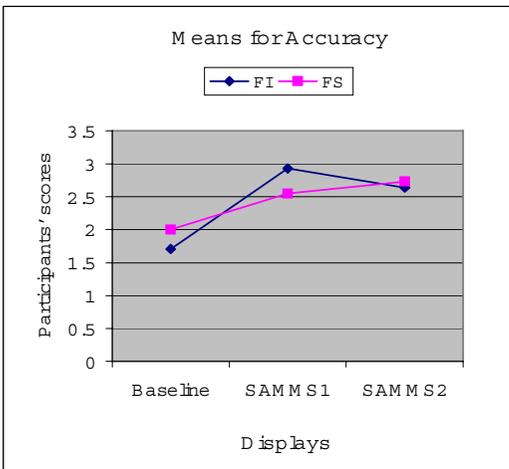
Figure 4. Accuracy means for experience levels and displays.

There was a significant difference in the participants' accuracy between the displays (F (2, 46) = 31.05, p < .05). There also was a significant interaction between the displays and experience levels (F (2, 46) = 3.75, p < .03) for the accuracy data (see Figure 4). Further post hoc paired comparisons were made to determine differences between the cell means given the significant interaction. The analysis required an adjustment for the nine simple effect comparisons by computing unique mean square error terms when comparing the displays. A Bonferroni type correction was used to determine which of the nine-paired comparisons were significant. This correction used a new alpha level that was obtained by dividing the original alpha level by the number of comparisons (.05/9 = .00556). Results of that analysis indicated that the baseline differed significantly from SAMMS1 and SAMMS2 among flight instructors (the baseline has a lower accuracy score). The baseline also was significantly different from

SAMMS2 among flight students (again the baseline has a lower accuracy score) while no other paired comparisons were significant. It is interesting to note that whereas students and instructors perform similarly on SAMMS2, instructors perform better than students on SAMMS1 do (see Figure 4). It would seem that instructors responded well to the organized layout of SAMMS1 and students responded well to the simplicity of SAMMS2.

There were significant differences in the accuracy data between the baseline and the two SAMMS prototypes. SAMMS1 and SAMMS2 provided the participant with the exact cause of the fault scenarios. The participant had only to click the appropriate button, which was clearly indicated on the display. The baseline, on the other hand, only provided instrument readings, which had to be interpreted. The baseline was a representation of the cockpit instrumentation system of a Piper Warrior. There was more than one instrument to be checked, and therefore, various readings to interpret. If something was missed, then an incorrect conclusion may have been drawn. Having the two SAMMS prototypes as supplemental systems significantly decreased the amount of false diagnoses. It also should be noted that the mean accuracy for SAMMS1 was the highest.

The results from the NASA TLX indicated the amount of workload experienced by the participants while using each display. Figure 5 shows the mean workload ratings (mental demand, effort, frustration, and performance) for the baseline, SAMMS1, and SAMMS2. This figure illustrates that participants experienced the higher levels of mental demand, effort, and frustration when they used the baseline than with SAMMS1 and SAMMS2. It also illustrates that participants felt they performed better with SAMMS1 and SAMMS2 than with the baseline.
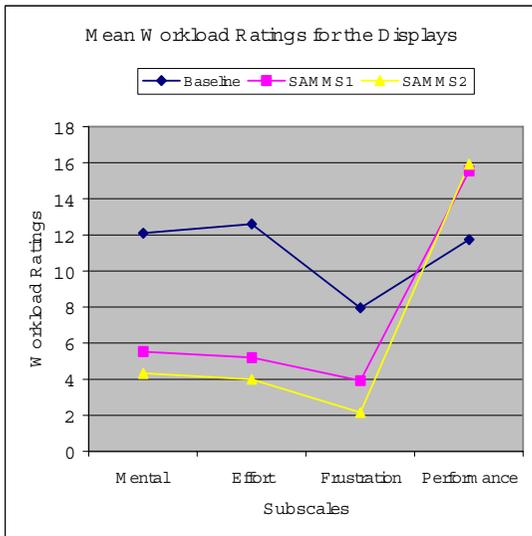
Figure 5. Mean workload ratings for the displays obtained from the NASA TLX.

The means of the workload scores for the baseline, SAMMS1 and SAMMS2 were 11.2 (SD = 2.39), 7.58 (SD = 2.22), and 6.61 (SD = 2.00), respectively. A single factor ANOVA was used to analyze the overall workload scores (alpha level = .05). There was a significant difference between the displays (F (2,72) = 29.91, p = .001). Further post hoc paired comparisons revealed that there was a significant difference between the baseline and SAMMS1 (SAMMS1 had the lower workload score) and between the baseline and SAMMS2 (SAMMS2 had to lower workload score). There was no significant difference between SAMMS1 and SAMMS2.
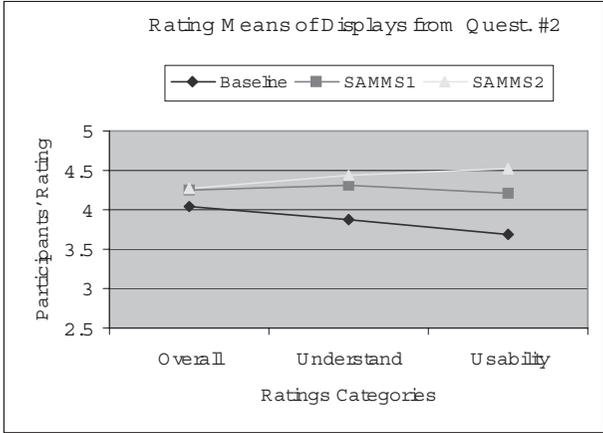
Figure 6. Rating means for each display obtained from Questionnaire #2.

In Questionnaire #2, the participants were asked to rate each display in three categories (understandability, usability, and an overall rating). The rating means for each display in the three categories are shown in Figure 6. The baseline has the lowest ratings while SAMMS2 has the highest rating.

The ratings for the three categories were averaged to give each participant a score. The scores were analyzed using a single factor ANOVA (alpha level = .05). The means and standard deviations for the baseline, SAMMS1, and SAMMS2 were 3.87 (SD = 0.71), 4.25 (SD = 0.85, and 4.41 (SD = 0.52), respectively. There was a significant difference between the displays (F (2, 72) = 3.93, p = .02). Further post hoc paired comparison determined that there was no significant difference between the displays. Comparison between the baseline and SAMMS revealed a difference of .54 (*Tukey's HSD* = .58).

Even though participants felt that they experienced less workload, performed better, and preferred the SAMMS2, their performance (diagnosis time and accuracy) using SAMMS1 was better. Many participants said that they preferred the simplicity of SAMMS2. With this display, the participants had only to click the highlighted button on the main screen to find out the exact cause of a failure. With SAMMS1, the participants had to go from the main menu to a second menu, and then to a third window that stated the cause of the failure. Some

participants stated that they preferred SAMMS1 because the buttons were larger and more organized than the buttons on SAMMS2 (see Figure 3).

Limitations and Recommendations

Based on the results of this study, it is recommended that future research on this topic use a flight simulator that is fully integrated with a functional SAMMS. The display comparison program was limited because it did not demonstrate all the capabilities of the SAMMS. Therefore, some of the beneficial features of the SAMMS, such as prognosis, could not be tested. Prognosis would have permitted participants to assess time to failure or at least provided a measure of criticality.Another untested feature of the SAMMS is de-cluttered mode, which gives the pilot the capability to customize the amount and types of information presented. With the SAMMS interface, the pilot will have the ability to turn "ON" or "OFF" the sound function or select the "min" option, in which only the cause of an impending failure is displayed. In addition, the emergency procedure checklist for the aircraft can be integrated into the SAMMS interface and displayed to the pilot during a failure. This additional feature of SAMMS would provide an electronic means to rapidly display critical emergency procedures that may not be recalled given the stress experienced during the emergency.

It also is recommended that future research consist of a sample that better represents the GA population of pilots. The selection of participants was based on convenience and not randomly selected. The pilots in the display comparison experiment were all flight instructors and flight students affiliated with a college flight program. Pilots that have not taken a class in the past two years or pilots that fly only occasionally were not represented.

Discussion

The more effective interface for displaying information presented by SAMMS to the GA pilot was assessed by four factors (diagnosis time, accuracy, workload, and participant rating). Diagnosis time and accuracy, which were obtained from the display comparison experiment, are objective measurements. Workload and participant rating, which were obtained from the NASA TLX and a questionnaire, are subjective measurements. Even though participants felt that they experienced less workload, performed better, and gave a higher rating to SAMMS2, their performances (diagnosis time and accuracy) were better when they used SAMMS1. In this study, the performance data (diagnosis time and accuracy) was used as the deciding factor.Therefore, SAMMS1 is considered the more effective of the two SAMMS prototype interfaces. The results of this study also indicated that SAMMS significantly improved accuracy of fault diagnoses while decreasing diagnosis time (though not significantly), and with relatively low levels of workload.

These results are similar to the results of the Navy's research. Like the Navy's HINTS, the SAMMS information is considered very useful as indicated by the

participants and as observed in the decrease in diagnosis time and the increase in accuracy.The SAMMS also reduced overall workload. The overall tone of the responses from participants was very positive with respect to the utility and usefulness of SAMMS.

References

Deaton, J., Glenn, F., Popp, E., Barba, C., & Bowers, C. (1998). *Investigation of aircrew information requirements for mechanical fault warnings*. (Tech. Rep. 981123.9803). Lower Gwynedd, PA: CHI Systems, Inc.

Deaton, J., Glenn, F., & Popp, E. (1999). *Development of a HUMS Interface System (HINTS) for helicopters*. Phase II (Tech. Rep. 990423.9701). Lower Gwynedd, PA: CHI Systems, Inc.

Glenn, F., Deaton, J., Barba, C., & Popp, E. (2000). Aiding Navy helicopter aircrews in handling mechanical fault emergencies. In *Proceedings of the Nineteenth Digital Avionics System Conference Proceedings*, Philadelphia, PA: IEEE, 7-19.

Hart, S. G. & Staveland, L. E. (1988). NASA Task Load Index (Version 1.0) [Computer Software]. Amsterdam, The Netherlands: Elseveir Science B.V.

Marsh, G. (1996, February). The future of HUMS. *Avionics Magazine*, 22-27.

Ritchie, M. L. (1988) General aviation. In E.L. Wiener & D.C. Nagel (Eds.) *Human factors in aviation* (pp. 561-589). San Diego, CA: Academic Press, Inc.

SPSS Inc. (2002) Statistical package for the social sciences, (Version 10.1 for Windows) [Computer software]. Prentice Hall.

Turbill, A. (1999) *The typical general aviation aircraft*. National Aeronautics and Space Administration (Tech. Rep. CR-1999-209550). Hampton, VA: Langley Research Center.

W ang, W. (1998) *Visual Basic 6 for Dummies (Deluxe Compiler Kit)*. New York, N Y: Hungry Minds.
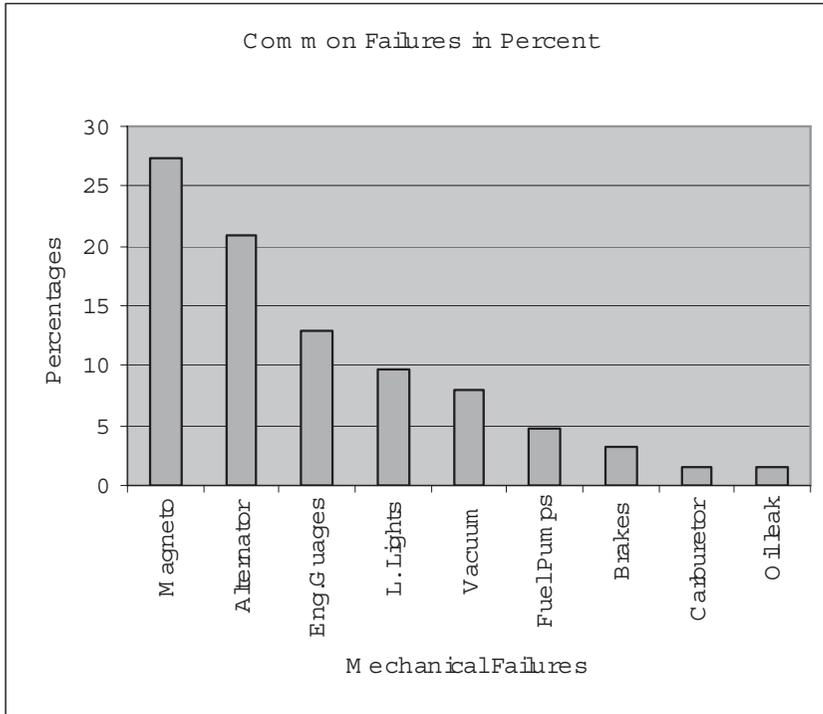
Figure A1. Most common failures experienced in-flight on the Piper Warrior in percentages.
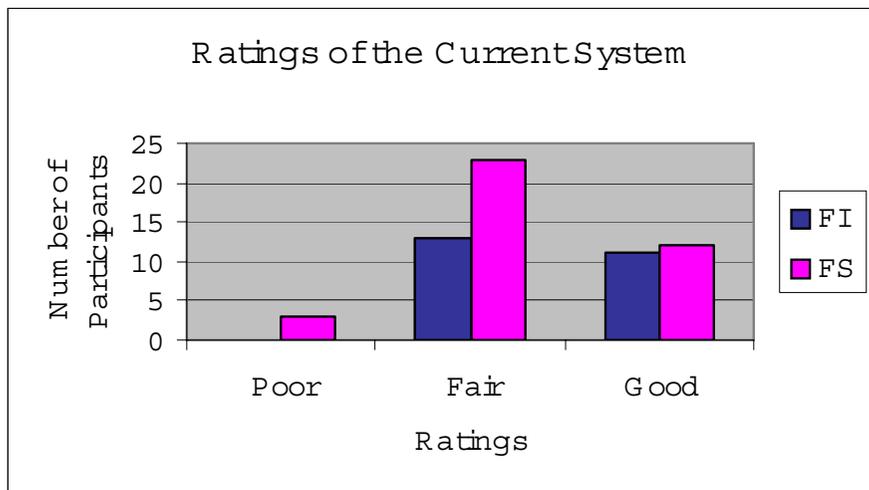
Figure A2. Participant's rating of the current fault detection system on the Piper Warrior.
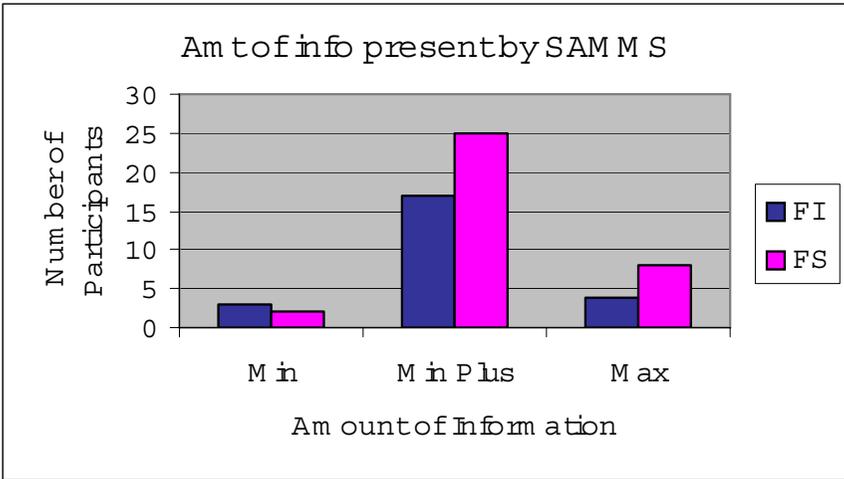
Figure A3. Amount of information participants indicated SAMMS should provide.

Table A1.

Comparison between the Features of the Baseline and SAMMS

Features	Baseline	SAMMS1&2	Suggested Features*
Instrument Readings	X		
Text		X	
Color		X	X
Advisory Warning	X	X	X
Cause of Failure		X	X
Affect on Aircraft Performance		X	X
Affect on other systems		X	X
EGT		X	X
Hydraulic Press. Gauge		X	
Ammeter	X	X	
Status of Carburetor		X	
Status of Magnetos		X	
Integrity of Flight Controls		X	X
Integrity of Landing Gear			X
Emergency Procedures			X

Note. The "X" indicates that the display has that particular feature.

* These are features participants would like to be incorporated into SAMMS.

The Effects of High-G Environments on Humans

Valerie J. Gawron
General Dynamics
150 North Airport Drive
Buffalo, New York 14225
716-631-6916
716-631-6990 (fax)

Abstract

The effects of high G on humans include: G-induced Loss of Consciousness (G-LOC), neck injury, vibration effects, reach envelope reduction, vestibular illusions, and inadvertent control activations. G-LOC has been experienced by about 20% of fighter pilots worldwide. Its physiology, symptoms, causal factors, and countermeasures, drawn from either in-flight incidents or centrifuge studies, are summarized in this paper. Neck injury, especially associated with placement of the head off center, is summarized. The effects of vibration on auditory and tactile perception are quantified and presented in figure format. Reduction in reach envelope also is quantified. Vestibular illusions and inadvertent control activations are described. A glossary and an extensive reference list are presented.

The Effects of High-G Environments on Humans

There are effects of high G on humans: G-induced Loss of Consciousness (G-LOC), neck injury, vibration effects, reach envelope reduction, vestibular illusions, and inadvertent control activations. All these are described in greater detail below.

G-LOC (+G-induced Loss of Consciousness)

G-LOC is not a new phenomenon (Kerr & Russell, 1944)[1], nor is it an uncommon phenomenon. G-LOC has been experienced by about 30% of all F-16 pilots (Pluta, 1984); about 14% of all United States (U. S.) Navy pilots (Johanson & Terry, 1986); about 17% of all Navy back/side seaters (Johanson & Pheeny, 1988); about 19% of all Royal Air Force (R. A. F.) pilots (Prior, 1987); and 10% of all Brazilian Air Force pilots (Alvim, 1995). The majority (63.3%) of reports of G-LOC in the R. A. F. involved pilots who were not controlling the aircraft at the time G-LOC was experienced (Prior, 1987). In the 1980s, two G-LOC episodes occurred in U. S. Air Force undergraduate pilot training every month (Whinnery, Glaister, & Burton, 1987).

The occurrence of G-LOC is directly related to brain oxygen levels. Oxygen is delivered to the brain via blood. The hydrostatic column of blood from the heart to eye level is approximately 30 cm long and weighs 22 mm Hg under 1 g conditions. When the heart cannot pump blood at a pressure of at least 22-mm Hg, G-LOC can occur. Since the brain maintains about a 5-second (s) oxygen reserve, pilots can go to very high-G levels without incurring G-LOC, but only if total time under high g is less than 5s (Gaines, 1987) or they are wearing blood pressure enhancing personal equipment. "During +Gz accelerations, blood is lost from the cortical areas of the brain first, followed by the central brain tissue, then ultimately the area of the brain stem. When +Gz accelerations cease or are reduced, the progression is reversed" (McCloskey, Tripp, Chelette, & Popper, 1992, p. 410). Transcutaneous Doppler ultrasonic flowmeter monitoring temporal arterial blood flow velocity has supported this finding (Rositano, 1980)[2].

There are three types of symptoms that occur with the exposure to G accelerations. The first is vision deficiencies. With moderate G onset rates, visual symptoms include tunnel vision and unequal peripheral light loss (Popper & Tripp, 1991), loss of peripheral vision (White, 1960[3]; Chambers & Hitchcock, 1963; Alvim, 1985; Albery, Jennings, Roark, Frazier, & Ratino, 1985[4]), as well as grey outs and/or blackouts (Yilmaz, Cetinguc, & Akin, 1999)[5]. These visual symptoms may be dependent on the preceding G experience as indicated in Table 1. Further, during maneuvers that result in the eyes being at different levels within +Gz field, vision is affected first in the eye located higher in the +Gz field (Whinnery, 1991).

---

[1] Centrifuge data from 5544 runs on 542 aircrew trainees and 13 subjects over 16 flights.
[2] Centrifuge data from 24 subjects with no previous centrifuge experience.
[3] Centrifuge data from 1 experienced subject exposed to 1 to 4 +G_z with an onset rate of 1 G/1.5 s.
[4] Centrifuge data from 14 male subjects exposed to 0.45 to 2.875 G_z/s.
[5] 95.7% of 325 Turkish jet pilots surveyed reported greyout and/or blackout.

Table 1

Subjects' reported incidents of visual light loss during 15-s exposure to +2.25 G_Z after exposure to varying levels of Gz

	+1 G_Z (control)	2 s -2 G_Z	5 s -2 G_Z	15 s -2 G_Z	Total
Light loss	0	1	2	4	7
No light loss	12	5	4	2	23
Total	12	6	6	6	30

(Banks, Grissett, Saunders, & Mateczun, 1995, p. 726)

Contrast required to detect differences increases in luminances increases with G (see Table 2). With high-G onset rates, there may be no visual symptoms prior to black out (Gillingham & Fosdick, 1988).

Table 2

Brightness Discrimination Thresholds

Direction of Acceleration	Level of Acceleration			
	Background Luminance in ft.L			
	31.20	2.90	0.29	0.03

(Braunstein & White, 1962, p. 932)

The second G-LOC symptom is incapacitation. Once G-LOC has occurred, the average incapacitation is 15 s but differs between types I and II G-LOC (see Table 3). Type I G-LOC is defined as "short direction and lack of convulsive activity, while Type II G-LOC is defined as "longer duration with convulsive activity" (Whinnery & Burton, 1987, p. 469). For type I G-LOC, the pilot is unconscious but not moving; for type II G-LOC, the pilot exhibits clonic movements, dreaming (Forster & Whinnery, 1988), and convulsions (Firth, 1993). Flailing behavior associated with Type II G-LOC may be associated with longer incapacitation times (see Table 4). Finally, the vast majority of G-LOC episodes are associated with stick release (Whinnery, 1986).

Table 3

Incapacitation Time as a Function of Type G-LOC

Type G-LOC	Absolute Incapacitation (s)	Relative Incapacitation (s)	Total Incapacitation (s)
I	14.9	13.3	28.2
II	20.4	17.3	37.3

(Whinnery & Burton, 1987) [6]

---

[6] Centrifuge data from healthy Air Force personnel.

Table 4

Incapacitation Time as a Function of Occurrence of Flail Behavior

	Flail	Non-Flail	Aircrew	Non-Aircrew
G-LOC	8.1	7.3	-	-
Absolute	20.4	14.9	-	-
Total (auditory)	37.3	28.2	31.7	29.3
Total (Visual)	36.6	28.8	29.3	36.3
Relative (auditory)	17.3	13.3	-	-
Relative (visual)	16.6	14.0	-	-

(Whinnery, Burton, Boll, & Eddy, 1987, pp. 634-635)

The third symptom is retrograde amnesia. Specifically, following incapacitation, retrograde amnesia occurs for another 10 s. Burton (1988) reported that 50% of subjects do not remember the G-LOC event.

Of great interest to the military aviation community is the assessment of the effect of G-LOC on flying performance. To assess this effect, a number of studies were conducted in centrifuges with laboratory tasks being used as surrogate flight tasks. The laboratory task that frequently studied was tracking because it is a surrogate for manual control of the aircraft. Tracking error increases with higher exposures to G. For example, Frazier, Repperger, Toth, and Skowronski (1982)[7] reported that tracking error was significantly greater at +5 Gz than at +1 Gz. The error was exacerbated when Gy was +/- 1 or +/- 2. This significant difference may be due to decreased pilot gain and decreased open-loop system crossover frequency, i.e., "getting behind the aircraft" (Sadoff & Dolkas, 1967)[8]. There is also an increase in throttle pointing bias (pitch down) from 4 to 6 +Gz. The bias is greater for rapid movements than for steady-state movements (Van Patten, Repperger, Hudson, & Frazier, 1983)[9].

The tracking error may be due to central hypoxia and G loading on the vestibular system (Cheung & Hofer, 1999)[10].

In addition to tracking, detection tasks have been used as surrogates for flying tasks related to target and threat detection as well as detecting changes in aircraft state. Reaction time to either light or sound stimuli significantly increases as +Gz increases (see Table 5). Other perceptual tasks have been used as well. For example, Frazier, Repperger, and Popper (1990) reported that perceived duration is significantly shorter than actual duration of exposure to +8 Gz for durations equal to or greater than 16 seconds. Also the ability to discriminate mass (105, 110, 115, 120, and 125 g) degrades as Gz increases from 4 Gz (Darwood, Repperger, & Goodyear, 1991)[11].

---

[7] Centrifuge data from 8 Air Force personnel.

[8] Centrifuge data from four human subjects.

[9] Centrifuge data from AFTI/F-16 project test pilots subjected to 1 to 6 +G_z

[10] Centrifuge data from 2 female and 9 male subjects.

[11] Centrifuge data from 9 male and 1 female Air Force officer

Table 5
Reaction Time to Light[12]

G-Level	Light		Sound	
	X	SD	X	SD
1	1.229	0.107	1.006	0.134
3	1.289	0.125	1.061	0.123
5	1.327	0.106	1.131	0.163

Cognitive task performance also is degraded as a function of +Gz (Chambers, 1963; Chambers & Hitchcock, 1963; Rogers, Ashare, Smiles, Frazier, Skowronski, & Holden, 1973; Piranian, 1974[13]; Whinnery & Staffstall, 1979; Albery, 1988[14]). However, advanced protection systems enable pilots to maintain cognitive functioning[15].

Finally, performance decrements may linger long after the G exposure. G-LOC aftereffects have included inability to takeoff, perform coordinated turns, and land[16]. In addition, Forster and Cammarota (1993)[17] reported significantly longer times to trim an aircraft after G-LOC (11.5 s) than before (7.7 s). However, time to acquire an airborne target was not significantly different before G-LOC (67.5 s) and one minute after G-LOC (66.3 s), yet Paul (1996) reported performance decrements in a flight simulator after +G exposure in a centrifuge in only one pilot. This pilot stalled and spun on takeoff. Post G-LOC symptoms are presented in Table 6. The data are from Brazilian Air Force pilots' questionnaire responses.

Table 6
Post-G-LOC Symptoms

Symptoms	N Reports from 193 Pilots
Dream-like State	5
Dream-like State & Psychomotor Incoordination	10
Dream-like State & Numbness	1
Psychomotor Incoordination	1
Scotoma	2
Scotoma & Tunnel Vision & Impaired Hearing	1

(Alvim, 1995)[18]

---

[12] Centrifuge data from 16 students.
[13] Centrifuge data from 11 naval test pilots exposed to up to +5 G_z
[14] Centrifuge data from 9 male subjects exposed to 2.75 or 3.75 G_z for 60 seconds.
[15] Centrifuge data from volunteers exposed to alternating +5Gz to +9Gz peaks with 5 sec plateaus until peripheral light loss or exhaustion.
[16] Simulator data immediately after exposure to GOR (0.1 G/sec), ROR (max onset from 1.4 G), 6 G with G-suit inflation or 5 G without, 8 G with or 7 G without, Simulated Air Combat Maneuver with 7, 8, and 9 G peaks over 90 seconds. 17 or 29 pilots did not experience G-LOC and did not show performance decrements. 12 of 29 experienced G-LOC – 11 of these did not show performance decrements.
[17] Centrifuge data from 7 healthy male naval personnel.
[18] Questionnaire data from 193 Brazilian Air Force pilots.

Physiological symptoms are listed in Table 7. In addition to these symptoms, Berkhout, O'Donnell, and Leverett (1973) reported excessive electrical activity in the areas of the brain associated with muscle activation[19]. Tachibana, Akamatsu, Nakamura, and Yagura (1994) suggested that +Gz induced autonomic imbalance may result in arrhythmia. Park, Seul, Park, Kim, and Cho (1994) reported a significant reduction in plasma atrial natriuretic peptide (ANP) and significant increases in renin concentration, heart rate, and blood pressure after exposure to +6 Gz for 30s[20].

Table 7

Number and Type of +G$_Z$-Induced Symptoms

Number of Occurrences	Symptom
16	abdominal pain
16	arm pain
1	clonic movements
5	disorientation, vertigo
3	hyperventilation
67	loss of consciousness
2	loss of consciousness with severe convulsion
15	neck pain
16	petechial hemorrhages
2	scrotal hematoma/discomfort

(Whinnery & Gillingham, 1983)[21]

Hamalainen, Toivakka-Hamalainen, and Kuronen (1999) reported two cases in which +Gz may have caused degenerative spinal stenosis of the cervical spine. Examination of X-rays of the spine of pilots in the Royal Netherlands Air Force suggested, "that frequent exposure to high +Gz forces might cause premature degeneration of the spine of F-16 pilots" (Henriksen & Holewijn, 1999, 1057). However, in response to a survey, pilots flying high performance jets (F-15, F-16) did not report higher incidence of chronic spinal symptoms or neck disease than did aircrews flying nonhigh performance jets (C-9, C-20, C-21, C-121, C-130, C-141) (Drew, 2000). Earlier research reported hypoxemia (Barr, 1962).

Five flight-envelope parameters determine the onset, duration, and recovery from GLOC: rate of onset, the magnitude of the +Gz, the offset rate, number of

---

[19] Centrifuge data from 8 subjects exposed to 6 45-second +G$_z$ exposures with peak values of 4.5 and +G$_z$ within 15 minutes.

[20] Centrifuge data from 7 men and 3 women exposed to up to 8 +G$_z$

[21] Centrifuge data over a three-year period at the Air Force School of Aerospace Medicine based on 544 subjects and 2,066 separate +G$_z$ exposures.

exposures, and preceding exposures. There also are eight non-flight envelope parameters that affect G-LOC: expectation, individual tolerance to +Gz, eye-heart vertical distance, age, anti-G straining maneuver (AGSM), PPB, G suits, and head position. One anti-G-LOC technique, use of drugs, has been suggested (Lambert & Wood, 1946) but not implemented.

Rate of Onset of the +Gz

The faster the rate of +Gz onset, the shorter the time to grayout (see Figure 1). The more gradual the onset of +Gz, the longer the incapacitation period (Whinnery, 1988; also see Table 8 for incapacitation times without countermeasures). A gradual +Gz onset rate results in an increased tolerance of approximately 1G as compared to a rapid onset rate (Burton, 1986) with an average difference of 1.9 +/- 0.7G (Edelberg, Henry, Maciolek, Salzman, & Zuidema, 1956)[22]. Hrebien (1983) developed the following equation to describe +Gz tolerance as a function of onset rate: +Gz tolerance = 4.05 + 3.7 (G onset rate)[23]. For example, with an on onset rate of +0.1 Gz /s, + Gz tolerance = 4.05 + (3.7 x 0.1) = 7.75 seconds. This equation has been expanded into a predictive model (Moore, Jaron, Hrebien, & Bender, 1993). More recently, Yilmaz, Centinguc, and Akin (1999) concluded "G-LOC seems to be a more common problem for pilots who fly rapid onset rate aircraft than pilots who fly high "G" capable but lower G onset rate aircraft" (p.709)[24].
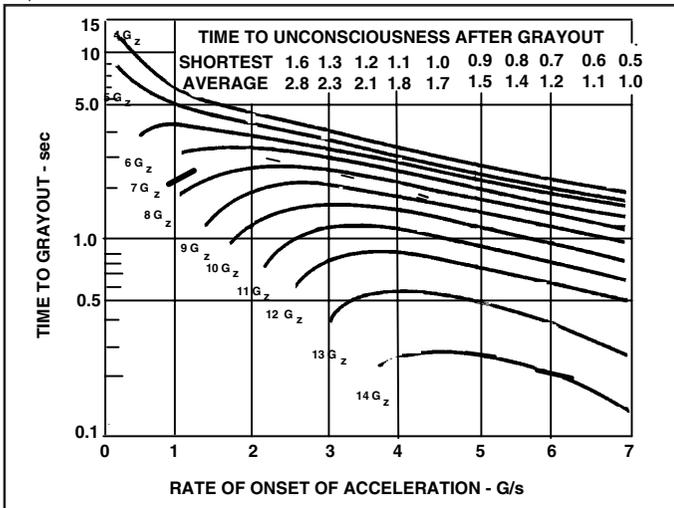
Figure 1. Time to Grayout as a Function of G Onset Rate. (Webb, 1964, p. 35)

---

[22] Centrifuge data from 32 human subjects without an anti-G suit in 49 series of runs.
[23] Centrifuge data from 5 human subjects exposed to ramp onset profiles ranging from 0.1 to 0.5 G/s.
[24] Survey data from 325 Turkish jet pilots and centrifuge data from 486 F-16, 801 F-4, and 256 F-5 pilots.

Table 8

Incapacitation Time as a Function of $+G_Z$ Onset Rate

Onset Rate $+G_Z$/s	G-LOC (s)	Absolute Incapacitation (s)	Relative Incapacitation (s)	Total Incapacitation (s)
0.08[13]	5.60 +/-1.35	12.36 +/-5.92	19.56 +/-13.66	31.92 +/-12.70
< 0.10[8]	4.0 to 5.5			
0.10[12]		19.30	15.90	34.90
0.10[13]	5.50 +/-2.01	7.67 +/-2.73	9.50 +/-4.97	17.17 +/-3.60
0.20[10]	4.30			
0.30[10]	3.00			
0.50[10]	3.30			
< 0.70[11]		11.00 to 15.00	17.50 to 22.00	34.00 to 36.50
> 1.00[12]		12.20	12.40	24.80
1.00[7]	5.40 +/-0.35			
1.40[10]	6.10			
1.60[11]		9.50 to 11.00	11.75 to 14.00	20.00 to 24.00
2.30[10]	8.00			
3.10[13]	8.24 +/-1.48	7.59 +/-3.14	5.40 +/-3.38	13.20 +/-4.36
3.70[10]	11.00			
3.79[13]	8.66 +/-0.63	10.47 +/-3.00	14.40 +/-10.05	25.04 +/-10.13
6.00[10]	15.00			
7.30[10]	16.50			

(Burton & Shaffstall, 1980[25]; Harding & Bomar, 1990[26]; Parkhurst, Leverett, & Shubrooks, 1972[27]; Stoll, 1956[28]; Whinnery, 1990[29]; Whinnery & Burton, 1987[30]; Whinnery & Whinnery, 1990[31])

The R. A. F. surveyed all R. A. F. pilots about their G-LOC experiences. The unclassified results indicated that the majority of G-LOC occurrences (i.e., 50.8%) were associated with a rapid G onset rate (2 to 4 Gz/s) (Prior, 1987). Surprisingly, only 37.0% of the G-LOC occurrences were associated with higher onset rates (4 to 6 Gz/s). There were G-LOC occurrences (12.2%) even with slow onset rates (1 to 2 Gz/s). These all occurred in flight.

---

[25] Centrifuge data from male Air Force personnel in a 30-degree seat.
[26] Centrifuge data.
[27] Centrifuge data from 46 male Air Force personnel aged 21 to 36 years old in a 13-degree seat with an anti-G suit.
[28] Centrifuge data from 300 experiments on 15 subjects.
[29] Centrifuge data from 500 G-LOC episodes.
[30] Centrifuge data from healthy Air Force personnel.
[31] Centrifuge data from 17 flight-qualified Navy and Marine fighter/attack aviators and 90 Air Force flight-qualified aircrew.

Forster (1994) reported two types of heart rate responses from 30 pilots in a centrifuge. For 43% of the subjects, heart rate gradually declined and then reached a steady rate 60s after exposure to +Gz. For the remaining 57% of the subjects, heart rate decreased. This decrease occurred 11s after maximum heart rate was attained. Martin, D'Aunno, Wood, and South (1999) reported that exposure to +Gz resulted in pulmonic insufficiency and tricuspid regurgitation[32].

Magnitude of the +Gz

The magnitudes of two +Gz levels are important: 1) maximum +Gz and 2) recovery +Gz. The higher the magnitude of maximum +Gz, the more severe the G-LOC symptoms are (see Figure 2 and Table 9). The higher the recovery +Gz, the longer the incapacitation period is (see Table 10). Higher recovery +Gz was studied as part of an aircraft auto-recovery system in which the aircraft accelerated to reduce the risk of surface to air missile strike. Humans have been able to sustain +12 Gz for three seconds in a centrifuge without G-LOC while wearing protective equipment (Cammarata, 1990).
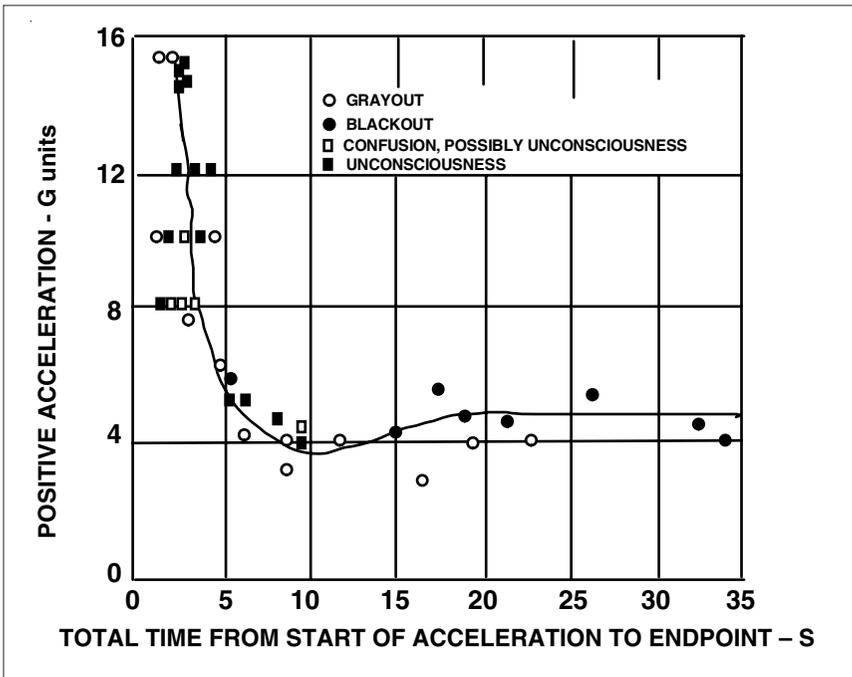
Figure 2. Severity of G-LOC Symptoms as a Function of Magnitude of +G$_z$ (Webb, 1964, p. 35)

---

[32] Echocardiographic data from 46 pilots and 201 nonpilots

Table 9

G-LOC Symptoms as a Function of Maximum +G_z

Symptom	Mean Threshold	Standard Deviation	Range
Grayout	4.1 +G_z[15]	+/- 0.7 +G_z[15]	2.2 to 7.1 +G_z[15]
Blackout	4.7 +G_z[15]	+/- 0.8 +G_z[15]	2.7 to 7.8 +G_z[15]
	5.4 +G_z[16]		4.9 to 6.2 +G_z[16]
Unconsciousness	5.4 +G_z[15]	+/- 0.9 +G_z[15]	3.0 to 8.4 +G_z[15]
	9.9 +G_z[14]	+/- 1.9 +G_z[13]	7.0 to 11.7 +G_z[16]

(Forster & Cammarota, 1993[33]; Gillingham & Krutz, 1974[34]; Parkhurst, Leverette, & Shubrooks, 1972[35])

Table 10

Incapacitation Time as a Function of Recovery +G_z

Recovery Level (+G_z)	Absolute Incapacitation (s)	Relative Incapacitation (s)	Total Incapacitation (s)
1.00	11.90 +/- 2.90	3.60 +/- 2.30	15.60 +/- 2.70
2.00	12.90 +/- 6.90	2.90 +/- 0.80	16.00 +/- 6.80

(Whinnery, Fischer, & Shapiro, 1989)[36]

Visual loss is typical with +7 Gz exposure (Gillingham, Makalous, & Tays, 1982[37]). Light loss may occur prior to G-LOC (see Table 11).

Table 11

Light Loss as a Function of +Gz

Statistic	80-degree light loss	23-degree light loss	central light loss
mean +G_z	4.2	4.5	5.3
range +G_z	2.7 to 5.7	2.9 to 6.4	3.6 to 7.0
standard deviation +G_z	0.7	0.8	0.8
mean duration (s)	5.4	5.1	6.8
range of duration (s)	1.9 to 17.0	1.9 to 11.9	2.1 to 23.4

(Webb, 1964, p. 43[38])

Offset Rate

The more gradual the offset of +Gz, the longer the incapacitation period is (see Table 12).

---

[33] Centrifuge data from seven male naval personnel.
[34] Centrifuge data from 1000 naval aviation cadets who were relaxed and unprotected.
[35] Centrifuge data from 46 male Air Force personnel aged 21 to 36 years old in a 13-degree seat with an anti-G suit.
[36] Centrifuge data from 8 males, 30° tilt-back seat, and no anti-G suit.
[37] Centrifuge data from 8 experienced centrifuge riders in F-4 profile.
[38] Centrifuge data.

Table 12

Incapacitation Time as a Function of $+G_z$ Offset Rate

Offset Rate $+G_z/s$	Absolute Incapacitation (s)	Relative Incapacitation (s)	Total Incapacitation (s)
0.97	10.47 +/- 3.00	14.40 +/- 10.05	25.04 +/- 10.13
2.75	7.59 +/- 3.14	5.40 +/- 3.38	13.20 +/- 4.36

(Whinnery & Whinnery, 1990[39]).

Number of Exposures

There is a tendency for the greater the number of exposures, the longer the incapacitation period (see Table 13). However, cardiovascular adaptation to + Gz does occur.[40]

Table 13

Incapacitation Time as a Function of Number of +Gz Exposures

Exposure	Absolute Incapacitation (s)	Relative Incapacitation (s)	Total Incapacitation (s)
1	7.80 +/- 3.11	7.40 +/- 2.79	15.20 +/- 4.87
2	11.60 +/- 4.51	4.40 +/- 1.82	16.00 +/- 3.16
3	11.80 +/- 3.35	12.00 +/- 12.47	23.80 +/- 12.76
4	11.50 +/- 3.54	4.00 +/- 1.41	15.50 +/- 2.12
5	12.00	8.00	20.00

(Whinnery & Jones, 1987[41])

Preceding –Gz Exposures

The immediate preceding exposure of less than +1 Gz may result in decreased +Gz tolerance. This phenomenon is called the "Push-Pull Effect" and has been named in at least one accident (Banks & Goodman, 1996). It has been estimated to occur in 11 to 67% of U.S. Air Force training flights (Michaud, Lyons, & Hansen, 1998)[42] and 12.5 to 29% of Air Force G-LOC accidents (Michaud & Lyons, 1998[43]). In a subsequent centrifuge study, heart rate was significantly greater during a +2.25 Gz exposure if preceded by 15 seconds of –2 Gz exposure than if preceded by 15 seconds at +1 Gz or 2, 5, 10, or 15 seconds –2 Gz (Goodman, Banks, Grissett, & Saunders, 2000).

Expectation

Unexpected exposures to +Gz result in longer incapacitation periods than expected exposures (see Table 14) but range from 14 to 31 s (see Table 15).

---

[39] Centrifuge data from 17 flight-qualified Navy and Marine fighter/attack aviators and 90 Air Force flight-qualified aircrew.

[40] In-flight data from 1 single-seat F/A-18 and 1 dual-seat F/A-18 (Newman and Callister, 1999).

[41] Centrifuge data from four human subjects.

[42] Head Up Display video from 240 Air Force air combat training engagements.

[43] Data from 24 Air Force G-LOC accidents between 1982 and 1996.

Table 14

Incapacitation Time as a Function of Expectation[44]

Expectation	Absolute Incapacitation (s)	Relative Incapacitation (s)	Total Incapacitation (s)
expected	12.00	12.00	24.00
liberate Not Self-Induced	12.00	12.00	24.00
liberate Self-Induced	12.00	3.50	16.00

(Whinnery & Burton, 1987[45])

Table 15

Incapacitation Time for Unexpected +G_Z-Induced Loss of Consciousness[46]

Incapacitation Period	Number of Subjects	Mean (s)	Max (s)	Min (s)
Absolute	55	16.6	30	5
Total (auditory)	52	31.0	63	10
Total (visual)	27	31.4	64	15
Relative (auditory)	52	14.5	51	1
Relative (visual)	27	14.9	50	4

(Whinnery, Burton, Boll, & Eddy, 1987, p. 633)

Individual Tolerance

For gradual G onset rates, four variables predict relaxed +Gz tolerance (Webb, Oakley, & Meeker, 1991): 1) height in cm (h), 2) weight in kg (w), 3) age in years (a), and 4) diastolic blood pressure (dbp): G tolerance = 10.610 - 0.051h + 0.029w + 0.022a + 0.012dbp.

> Well-trained pilots have a greater +Gz tolerance (Darrah & Klein, 1987):
> Well-trained pilots: +GzRMS[47] = 17.03T - 0.26
> Relaxed pilots: +GzRMS = 7.62T - 0.15

where +GzRMS is the root-mean squared +Gz time tolerance to fatigue and T is time in s. Low g tolerance is defined as inability to tolerate a 7-g, 15-s, rapid-onset G profile or an 8-g, 15-s, rapid onset G profile for an F-16 configured seat (i.e., 30-degree seatback angle with elevated rudder pedals) (Gillingham, 1987). About one percent of the actively flying Air Force aircrew has low g tolerance[48]. The 30-degree F-16 seat-back is associated with a 0.5 to 1.0 G of protection over the 13-degree seat for both relaxed and straining subjects (Gillingham, 1988). Weight training may extend simulated air combat maneuvering time up to 300 seconds (Bulbulian, Crisman, Thomas, & Meyer 1994). After 300 seconds there were no differences between weight-trained and non-weight-trained subjects.[49] The correlation between lower body negative pressure (LBNP) tolerance and +Gz tolerance increases as LBNP onset rate increases or as +Gz

---

44   Centrifuge data from male subjects.
45   Centrifuge data.
46   Centrifuge from male subjects
47   Multiple sequential exposures to +G_Z acceleration pulses.
48   Centrifuge data derived from over 700 Tactical Air Command pilots who received high-G centrifuge training (Gillingham, 1986).
49   Centrifuge data from 4 naval aviators.

onset rate decreases (Ludwig, Krock, Doerr, & Convertino, 1998)[50]. LBNP has provided effective protection against negative Gz (Beck & Tripp, 1989)[51].

The following factors also lower an individual's tolerance to +Gz exposure: 1) chronic and acute hypertension, 2) high temperature, 3) hypoxia, 4) hypoglycemia, 5) stress, 6) dehydration, 7) infection, 8) alcohol ingestion, 9) varicose veins, 10) hemorrhoids, 11) hernia, 12) high myopia, and 13) glaucoma (Gillingham & Krutz, 1974, p. 39). G-LOC mishap pilots have significantly higher systolic blood pressure and significantly less aircraft-specific flight hours than other Air Force pilots (Lyons, Harding, Freeman, & Oakley, 1992). Finally, one study has shown that women have a lower +Gz tolerance than men (see Table 16), while another reported no difference (Chelette, Albery, Esken, & Tripp, 1998)[52].

Table 16
+G$_z$ Tolerance[53] by Gender

		Women	Men
GOR1	0.067 +G$_z$/s	4.81 +/- 0.79	4.79 +/- 0.75
ROR	1.000 +G$_z$/s	3.39 +/- 0.55	3.35 +/- 0.49
GOR2		4.64 +/- 0.83	4.52 +/- 0.72
GORs		5.59 +/- 0.92	5.67 +/- 0.81

(Gillingham, Schade, Jackson, & Gilstrap, 1986[54])

G-LOC is associated with: 1) crewmember not flying, 2) disconnected G-suit hoses, 3) fatigue, 4) improper diet, 5) being unprepared, and 6) lack of physical conditioning (Pluta, 1984).

Burton (1980) stated "the person whose energy, metabolic, and cardiovascular states are least disturbed by high-G exposure is the person who will perform best and become least fatigued during repeated aerial combat maneuvers" (p. 1191). In a recent survey, fighter pilots perceive physical fitness as an important factor in operating in +Gz environments (Newman, White, & Callister, 1999)[55].

Eye-Heart Vertical Distance
The greater the eye-heart vertical distance, the lower the +Gz tolerance (Burton, 1986):

$$G = (Pa \times d)/h = (98.4 \times 13.6)/h = 7.235/h$$

---

[50] Centrifuge data from two experiments with a total of 17 male volunteers, 12 of whom participated in both experiments.
[51] Centrifuge data from 10 male subjects exposed to –100, -50, and 0 mm Hg LBNP during –1.0, -1.5, and –2 G$_z$
[52] Centrifuge data from 8 male and 8 female nonpilot Air Force airmen.
[53] Defined subjectively as 100% peripheral light loss or 50% reduction in central light intensity.
[54] Centrifuge data from 102 Air Force women and 139 Air Force men.
[55] Survey data from 42 Royal Australian Air Force pilots.

where G is +Gz, Pa is arterial blood pressure (the average for pilots is 98.4), d is the density of Hg (13.6), and h is the eye-heart vertical distance in cm. Eye-heart vertical distance can be decreased by increasing the seat-back angle.

Age

There is a tendency for relaxed G tolerance to increase with age for initial Gradual Onset Rate conditions (see Table 17).

Table 17

$+G_Z$ Tolerance by Age

		30-34	35-39	40-44	45-49	50-54
GOR1[56]	0.067 G/s	4.6 ± 0.6	4.6 ± 0.7	4.8 ± 0.8	5.2 ± 0.8	5.3 ± 0.8
ROR[57]	1.000 G/s	3.5 ± 0.3	3.6 ± 0.3	3.7 ± 0.3	3.6 ± 0.6	3.9 ± 0.5
GOR2	0.067 G/s	4.4 ± 0.9	4.4 ± 0.6	4.4 ± 0.4	4.7 ± 1.0	4.9 ± 0.8
N		11	10	5	4	4

(Hull, Wolthuis, Gillingham, McCracken, & Triebwasser, 1977[58], p. B14-2)

AGSM

AGSM increases +Gz tolerance (Burton, 1986):

$$G = (Pa + S) \times d/h = (98.4 + S) \times (13.6)/(33.4) = 40.05 \times 0.407S$$

where G is +Gz, Pa is arterial blood pressure (the average for pilots is 98.4), S is the effect of AGSM on +Gz tolerance (can be inferred from esophageal pressure, Pes; the average Pes for pilots and a 30-degree seat-back angle is 44.8), d is the density of Hg (13.6), and h is the eye-heart vertical distance in cm (the average for pilots and a 30degree seat-back angle is 33.4). For maximum effectiveness, each individual must choose the best AGSM (i.e., M-1, L-1, Q-G, Hook) (Whinnery & Murray, 1990). Common problems include: breathe too slow, inhale too long, do "not get the jump on the Gs", and talk during G exposures (Lyons, Marlowe, Michaud, & McGowan, 1997)[59]. In addition, a breathing system that meets the Air Standardization and Coordination Committee limit of +/- mmHG is critical (Whitley, 1997)[60]. Finally, G-suits with standard pressure (1.5 psig x G-1) are associated with significantly longer GOR (+0.1G.s-1) duration tolerance than those with lower pressure (1.1 psig x G-1) (Krock, Balldin, Harmo-Ringdahl, Singstad, Linder, & Siegborn, 1997)[61].

---

[56] Gradual Onset Run, +G_Z force applied at a rate of 0.067 +G_Z/s to a possible maximum of 6 +G_Z.

[57] Rapid Onset Run, +G_Z force applied to a rate of 1.0 +G_Z/s to progressively higher +G_Z levels and maintained for 15s at each level.

[58] Centrifuge data from healthy aircrew, seat-back angle of 13o, no anti-G suit.

[59] Questionnaire data from 40 F-16 and 46 F-15 U.S.A.F. pilots.

[60] Centrifuge data from 61 F/A-18 pilots, seat-back angle 15°, anti-G suit.

[61] Centrifuge data from 6 male subjects exposed to a simple profile of +7 G_Z sustained for 15 s and a complex profile of +4.8 G_Z for 10s followed by +2 G_Z for 10s.

PPB

PPB increases +Gz tolerance (Leverett at al., 1973[62]; see Table 18).

$$G = [(S \times d) + k]/h = [(S \times 13.6) + 1338)]/33.4 = (13.6S + 1338)/33.4$$
(Burton, 1986)

where G is +Gz, S is the effect of PPB on +Gz tolerance, d is the density of Hg (13.6), k is a constant of 1338, and h is the eye-heart vertical distance in cm (the average for pilots and a 30degree seat-back angle is 33.4). PPB can be augmented with anti-G inflation trousers and a chest counter-pressure waistcoat (Prior, 1991)[63] or extended G-suits (Goodman, Fraser,Ackles, Mohn, & Pecaric, 1993). PPB has only two side effects: 1) presence of a dry cough (Travis & Morgan, 1994) and 2) fall in the left ventricular preload which may be mitigated by full-coverage anti-G suits (Goodman, de Yang, Kelso, & Liu, 1995)[64].

Table 18
+G_Z Tolerance With and Without PPB

	Assisted PPB	
	Without	With
Mean	5.1	5.5
Standard Error	±0.1	±0.2

(Burns, 1988[65])

PPB exacerbates arm pain during high + Gz in aircraft in which the stick and throttle are located 20 cm or more below the pilot's heart level. Arm arterial occlusion cuffs (AAOC) do not alleviate arm pain (Green, 1997)[66] nor do other arm pressure covers (Watkins, Welch, Whitely, & Forster, 1998)[67]. Linde and Balldin (1998) suggested that presence of arm pain should be considered during design of either aircraft or protective equipment[68].

Balldin, Tong, Marshall, and Regna (1999) tested the hypothesis that positive pressure breathing (PPB) during G in combination with extended coverage anti-G suits (ECGS) would increase the risk of premature ventricular contractions (PVCs). PVCs did not occur during gradual onset rate exposures but did occur during both simulated and tactical aerial combat maneuvers. However, there was no significant difference between standard equipment and the PPB/ECGS combination[69].

---

[62] Centrifuge data from 12 male subjects exposed to 2 series of 60 sec exposures to 3, 6, and 8 G followed by 60 sec at 1 G.
[63] Centrifuge data from 4 experienced centrifuge subjects exposed to 3 to 7 +G_z
[64] Centrifuge data from 9 human subjects with anti-G suit.
[65] Centrifuge data from 5 relaxed human subjects with anti-G suit.
[66] Centrifuge data from 12 male Swedish Air Force fighter pilots.
[67] Centrifuge data from 7 male volunteers.
[68] Questionnaire data from 35 Swedish fighter pilots.
[69] Centrifuge data from 2 female and 12 male volunteers.

G-Suit

G-suits provide protection against G-LOC. Not all G-suits provide the same amount of protection, however. Forster, Cammarota, and Whinnery (1994) compared incapacitation associated with a standard CSU15-P suit versus the same suit inflated to 10 psi immediately upon G-LOC and Burton (1988) evaluated the CSU-BA/P (see Table 19). Meeker (1991) reported that the Advanced Technology Anti-G Suit (ATAGS) enabled reduction from 10 psig to 8 psig at 9 G without decreasing G protection[70]. The Swedish Tactical Flight Combat Suit (TFCS) caused no abdominal pain associated with the inflation of the abdominal bladder. This was not the case with the Air Force standard G Suit (Balldin, Krock, Danielsson, & Johansson, 1996)[71]. Further, some suits such as the USAF's Advanced Technology Enhanced Design G-Ensemble (COMBAT EDGE) have integrated a PPB system with the G-suit (Nunneley, French, Vanderbeek, & Stranges, 1995) and have no associated G-LOC events (Tong, Balldin, Hill, & Dooley, 1998)[72]. Suits that include PPB provide enhanced G-LOC protection (Albery & Chelette, 1998)[73]. G-suits with increased body coverage provide better G protection than G trousers only (Albery, 1997)[74]. In addition, G-suit inflation may increase cardiovascular responsiveness to carotid baroreceptor stimulation (Convertino & Reister, 2000)[75].

Table 19

+Gz Tolerance With and Without Inflation

	No Inflation	Inflation
GOR (0.1 G/s)[76]	6.1 ± 2.7	8.9 ± 4.0
ROR (2-4 s to max G)	5.2 ± 2.0	6.9 ± 2.0
GOR (0.1 G/s)[77]	4.1	4.9
ROR (0.6 G/s)	3.8	4.7

Head Position

Bringing the head lower by increasing the seatback angle from 30?? to 65?? resulted in at least a 30% increase in relax +Gz tolerance (Burns & Whinnery, 1984)[78]. Significant differences in peak +Gz occur with 0.1 G/s onset, relaxed, and no G-suit inflation (Tong, Hill, Tripp, & Webb, 1994)[79]. Specifically, higher +Gz were associated with over the shoulder (right Gz ~ 6.75, left ~ 6.4) than straight-ahead (~ 5.0) or above (~ 5.4) positions. No differences were found at 0.5 G/s onset straining and G-suit inflated.

---

[70] Centrifuge data from 6 male volunteers.
[71] Centrifuge data from 4 Swedish test pilots.
[72] Centrifuge data from 3 female and 12 male volunteers.
[73] Centrifuge data from 6 subjects.
[74] Questionnaire data from 42 Royal Australian Air Force F/A-18 pilots.
[75] Physiological data from 12 men (not pilots) in a lower body negative pressure chamber with and without G-suit.
[76] Centrifuge data from 81 naval aircrew members receiving high-G centrifuge training.
[77] Centrifuge data from 10 males from an acceleration subject pool.
[78] Centrifuge data from 7 or 9 subjects for each seat configuration.
[79] Centrifuge data from 12 human subjects with centrifuge experience.

Fighter pilots have developed head positioning strategies including (from most to least frequent): set head position before application of G, move as required, brace head against ejection seat structure, move head/neck in only one plane at a time, use shoulders to aid rotation of head, keep head aligned with body under G, restrict movement under high G, move only under low G, move upper body as well as neck/head, and brace head against aircraft canopy (Newman, 1997)[80].

Neck Injury

Rapid G onset may cause neck injury. Such injury is especially possible whenever the pilot's head is off center from the spine or the harness is loose (personal communication, Drs. Kent Gillingham & James Whinnery). Pilots have impaled their helmet visors on the top of the centerstick after rapid onset G. Short-duration (i.e., less than 1 s), high accelerations (i.e., greater than 25 g) can cause severe injury (see Figure 3 parts a and b and Figure 4 parts a and b). Frequent muscle endurance training may reduce the occurrence of neck pain (Hamalainen, Vanharanta, & Bloigu, 1993)[81]. Frequent exposure to high + Gz forces may cause degeneration of the cervical intervertebral disks, especially in the C3-4 disk (Hamalainen, Vanharanta, & Kuusela, 1993)[82].
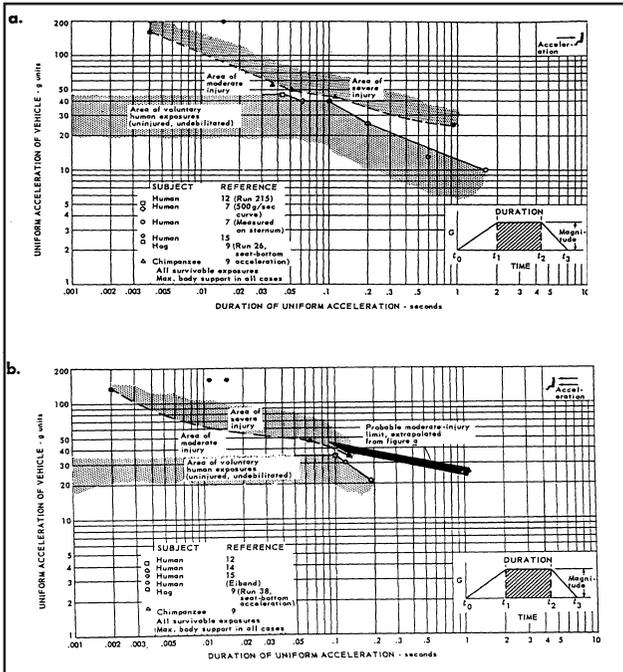
Figure 3. Part a is the Effect of $+G_x$ and Part b is the Effect of $-G_x$.

---

[80] Centrifuge data from 6 human subjects.
[81] Questionnaire data from 27 male student fighter pilots enrolled in the Finnish Air Force Academy.
[82] Based on resistive magnetic resonance scanner data of 12 male senior Finnish fighter pilots and 12 controls.

Figure 4. Part a is the Effect of $-G_z$ and Part b is the Effect of $+G_z$.

In a follow-up study, Hamalainen, Vanharanta, and Bloigu (1994) reported 37.9% incidence of acute in-flight neck pain. Incidence of pain was significantly correlated to the number of flight hours (+0.95). Buhrman and Perry (1994) reported that increases in helmet weight (3.2 to 6.6 lb.) and seat acceleration (up to 10 Gz for humans and +15 Gz for manikins) resulted in increased compression, shear, and rotational forces on the neck. In the Royal Australian Air Force, 44 of 52 fighter pilots surveyed reported neck injury under +Gz, usually simple muscle sprains (Newman, 1997). Weights greater than 4.5 lb. exceeded injury limits at +15 Gz[83]. Kikukawa, Tachibana, and Yagura (1995) surveyed 129 F-15 pilots. Of these, 89.1% reported muscle pain on different occasions especially in the checking six and forward bend positions. Hamalainen, Vanharanta, Hupli, Karhu, Kuronen, and Kinnunen (1996) reported twenty pilots in the Finnish Defence Forces experienced a 4.9 mm decrease in body height after maneuvering under high +Gz (+6.2 to +7.8 Gz) for 40 minutes. As indicated in Figures 3 and 4, the direction of the G force greatly affects the probabil-

[83] Centrifuge data from 14 human subjects (with centrifuge experience up to $+10G_z$) and the Hybrid III neck in the ADAM manikin (up to $+15 G_z$).

ity of injury. The direction also affects physiological tolerance. Physiological tolerance is highest for +Gx, next for - Gx, next for +Gz, and lowest for -Gz direction of force (Webb, 1964, p. 33; see Figure 5 a and b). "-3 Gz is ... the upper limit of human tolerance" (Gillingham & Krutz, 1974, p. 53). Burton et al. (1999) stated that there is "a direct relationship between degenerative diseases of the spine and repeated exposures to sustained G.

a.
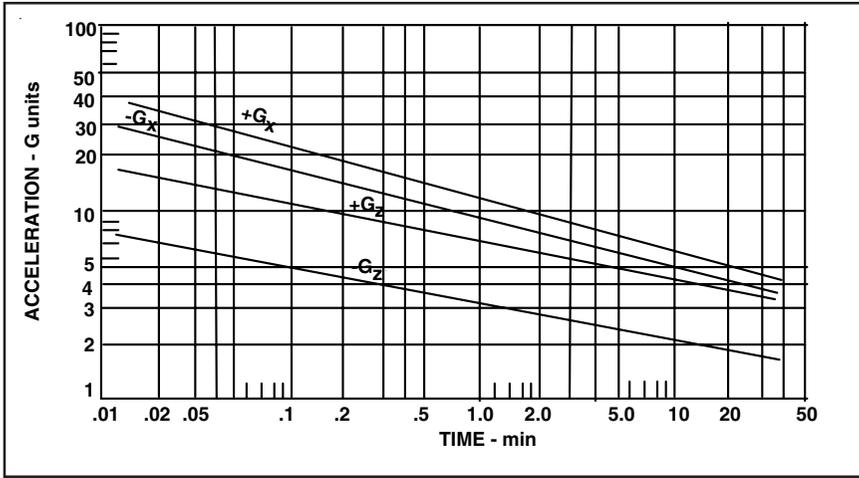
b.
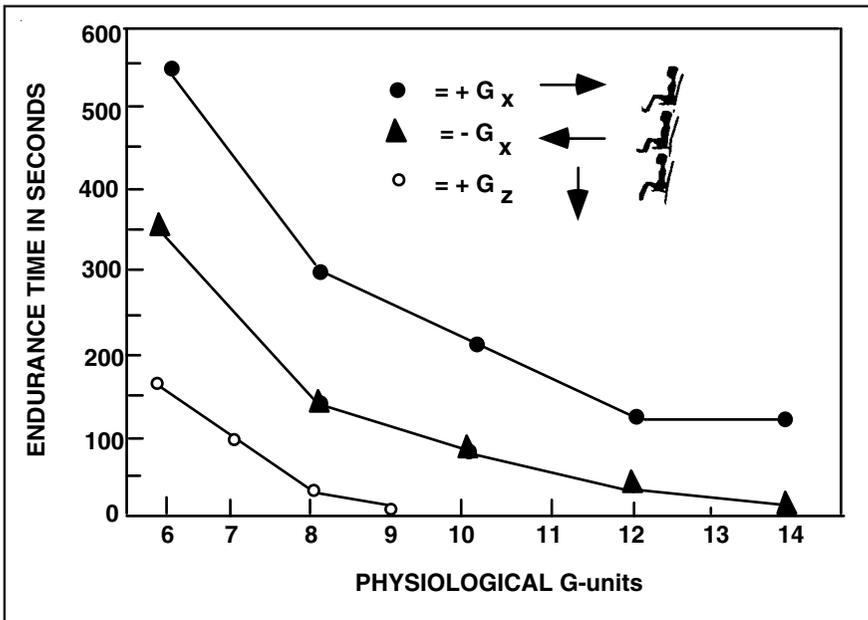
Figure 5. Human endurance time as a function of direction and magnitude of g. (Webb, 1964, p. 36)

Vibration affects both auditory and tactile perception. Low frequency vibration induces decreased tactile sensitivity, motion sickness, and difficulty in focusing (see Figure 6).

Figure 6. Effects of vibration. Guignard & Irving (1960)

Decrements in human performance are summarized in Table 20.

Table 20

Summary of decrements in human performance

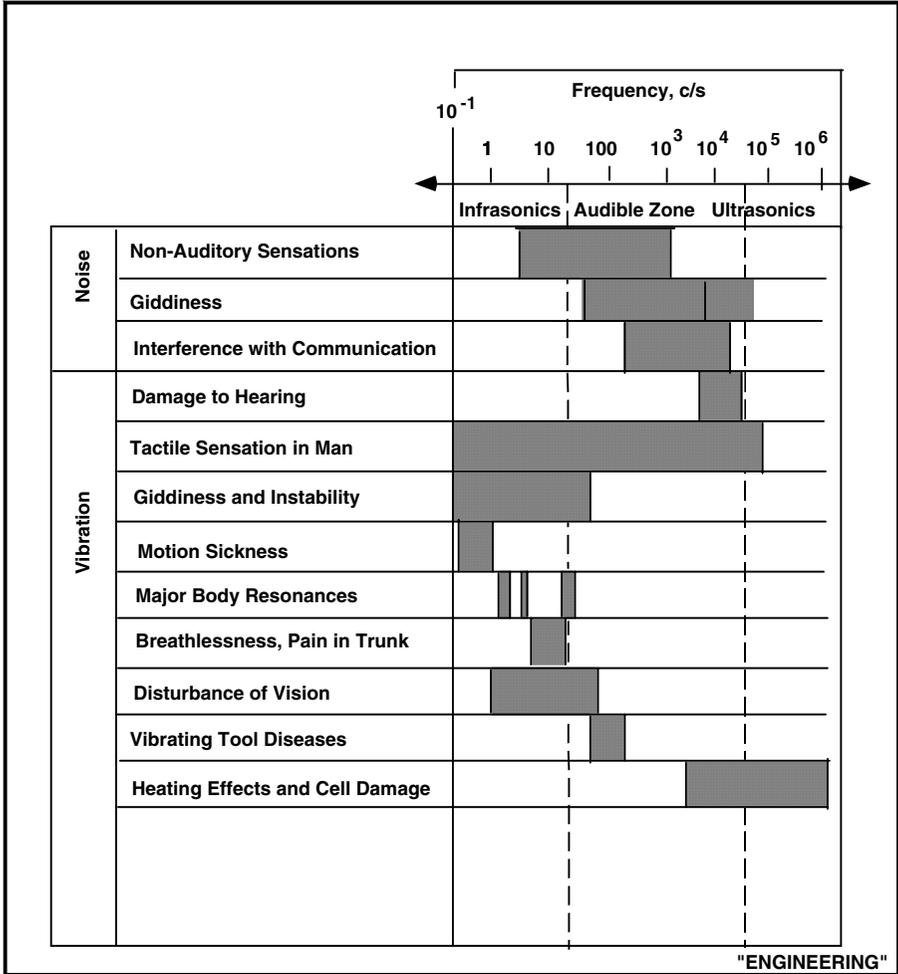| Category	Vibration Characteristics	Type of Decrement	References
Vision:	Whole-body vibration 8 to 9 Hz	Detection of visual blur	Griffin (1975)
	Display vibration, less than 3.5 Hz	Loss of legibility	Moseley & Griffin (1986)
	Whole-body vibration, 5.6 to 11.2 Hz	Loss of legibility	Moseley & Griffin (1986)
	Whole-body vibration, 4 to 11.2 Hz	Larger display characters lessened vibration effects on legibility	Lewis & Griffin (1979)
	Whole-body vibration, 3.15, 4, and 5 Hz	Increased character spacing lessened vibration effects on legibility	Moseley & Griffin (1986)
	Whole-body vibration, 2, 4, 6, 8, and 10 Hz	Viewing a display at infinity lessened vibration effects on reading ability	Wilson (1974)
Memory and central processing:	Severe whole-body vibration, 70 Hz	Decrements in continuous counting performance	Ioseliani (1967)
	Whole-body vibration, below 20 Hz	No decrements in mental addition, pattern recognition or navigational behavior	Hornick (1973)
	Whole-body vibration, 3 to 8 Hz	No effects on counting tasks; decrements in reading task as time increased	Shoenberger & Harris (1981)
Manual control:	Whole-body vibration, 3 to 8 Hz	Tracking decrements by virtue of arm, hand, and finger displacement	Allen et al. (1973) Lewis & Griffin (1978) Schmitz et al. (1960)
	Whole-body vibration, 2, 6, and 10 Hz	Side-mounted stick superior at 2 Hz; center-mounted stick superior at 10 Hz	Shoenberger & Wilburn (1972)
	Whole-body vibration 0.17 Hz	Keypunching unaffected; tracking severely degraded	McLeod et al. (1980)
	Whole-body vibration, 0.25 Hz	Degraded navigational and tracking performance	Wiker et al. (1980)

Reach Envelope

Schafer and Bagian (1993) reported a 17% reduction in forward reach when subjects were exposed to +4 Gx. There was also an 8% reduction in left overhead reach as compared to right overhead reach when exposed to +4 Gx.

Vestibular Illusions

Translational and angular accelerations may induce one or two vestibular illusions. The first is the G-excess illusion "where the pilot is exposed to an extended duration acceleration magnitude greater than 1G, movements of the head can create erroneous perception of both orientation and angular motion" (Pancratz, Bomar, & Radden, 1994, p. 1131). The second is the cross-coupling

illusion, "after prolonged exposure to a constant velocity rotation, a crewmember's semicircular canals adapt to the angular motion and he or she receives no further vestibular cue of the motion" (p. 1131).

Control Activation

Orrick, York, and Cohen (1976) examined the effects of g on ejection seat activation. There were significant conditions, subject, and conditions by subject interaction. The means are presented in Table 21.

Table 21
Pull Forces by Condition[84]

Simulation	G_z	G_y	G_z	X	SD
Recovery from Dive	0	0	5	84.08	12.08
Cold Catapult Stroke	2.5	0	1	113.92	14.94
Level Flight	0	0	1	107.67	13.73
Braking Deceleration	-2.5	0	1	94.25	12.76
Recovery from Dive with Buffet	0	0	5*	87.45	14.21
Inverted Flight	0	0	-1	115.79	14.01
Spin	2.5**	0**	1.75*	90.35	17.62
Skid	-2.5	1.5	1	86.77	15.03

Definitions

AGSM	Anti-G straining maneuver
blackout	Loss of all vision (Newsom, Leverett, & Kirkland, 1968).
clonic movements	Rapid contractions and relaxations of muscles
grayout	Loss of peripheral vision (Gillingham & Krutz, 1974, p. 30)
g	Displacement acceleration
G	Physiological acceleration, the total reactive force divided by the body mass (Webb, 1964, p. 33)
G-LOC	$+G_z$-induced Loss Of Consciousness
$-G_x$	Gravitoinertial force acting back to chest, associated with aircraft decreasing forward velocity (e.g., application of speed brakes) or a steep dive (Gillingham & Krutz, 1974)
$+G_x$	Gravitoinertial force acting chest to back (Burton & Shaffstall, 1980), causes the heart to be displaced back toward the spine (Webb, 1964, p. 33), associated with the aircraft increasing forward velocity (e.g., application of afterburner) or a steep climb (Gillingham & Krutz, 1974).
$-G_y$	Gravitoinertial force acting to the left; associated with aircraft in left slip or left skid (Gillingham & Krutz, 1974).

---

[84] Centrifuge data from 16 male Navy personnel
* +/- 0.3G @ 10 Hz (buffet)
** +/- 0.25G @ 0.3 Hz (oscillation)

+G$_y$	Gravioinertial force acting to the right; causes the heart to be displaced to the left (Webb, 1964, p.33); associated with aircraft in right slip or right skid (Gillingham & Krutz, 1974).
-G$_z$	Gravioinertial force acting foot to head; associated with aircraft in inverted flight, push-over into dive, or "outside" maneuver (Gillingham & Krutz, 1974).
+G$_z$	Gravioinertial force acting head to foot (Parkhurst, Leverett, & Shubrooks, 1972); causes the heart to be displaced downward (Webb, 1964, p. 33); associated with aircraft in level flight, coordinated turn, pull-up from a dive, or "inside" maneuvers (Gillingham & Krutz, 1974).
hematoma	Swelling filled with blood.
petechial	Small hemorrhage in the skin or mucous membrane.
+R$_x$	Angular acceleration that causes the heart to rotate to the left, radians per second^2 about the x axis (Webb, 1964, p.33).
+R$_y$	Angular acceleration that causes the heart to pitch down, radians per second^2 about the y axis (Webb, 1964, p.33).
+R$_z$	Angular acceleration that causes the heart to yaw right, radians per second^2 about the z axis (Webb, 1964, p.33).
PPB	Positive Pressure Breathing
s	seconds
Scotoma	Gap in visual field

Acknowledgments

References

Albery, W. B. (January 1988). *The effects of biodynamic stress on workload in human operators (AAMRL-TR-88-004).* Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Albery, W. B. (1997). Evaluation of six G protection ensembles during 5 to 9G simulated aerial combat maneuvers, *SAFE Journal 27* (2): 92-99.

Albery, W. B. & Chelette, T. L. (1998). Effect of G suit type on cognitive performance. *Aviation, Space, and Environmental Medicine, 69* (5), 474-479.

Albery, W. B., Jennings, T., Roark, M., Frazier, J., & Ratino, D. (April 1985). *Simulation of high +G$_z$ onset in the Dynamic Environment Simulator (ARAMRL-TR-85-030).* Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Allen, W.A., Jex, M. R., & Magdaleno, R. E. (1973). *Manual control performance and dynamic response during sinusoidal vibration (AMRL-TR-73-78)*. Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

Alvim, K. M. (1995). Greyout, blackout, and G-loss of Consciousness in the Brazilian Air Force: A 1991-92 survey. *Aviation, Space, and Environmental Medicine 66*: 675-677.

Balldin, U. I., Krock, L. P., Danielsson, C. H., & Johansson, S. A. (1996). Centrifuge man-rating of a conceptual internal abdominal bladder restraint in an extended coverage anti-G suit. *Safe Journal, 26*(2): 42-46.

Banks, R. D. & Goodman, L. S. (April 1996). Neurological influence in push-pull effect. Proceedings of the Neurological Limitation of Aircraft Operations: Human Performance Implications Conference. AGARD-CP-579. 1-1 to 1-10.

Banks, R. D., Eng, B., Grissett, J. D., Saunders, P. L., & Mateczun, A. J. (1995). The effect of varying time at $-G_z$ on subsequent $+G_z$ physiological tolerance (push-pull effect). *Aviation, Space, and Environmental Medicine, 66*: 723-727.

Banks, R. D., Grissett, J. D., Turnipseed, G. T., Saunders, P. L., & Rupert, A. H. (1994). The "push-pull effect". *Aviation, Space, and Environmental Medicine, 65*: 699-704.

Barr, P. O. (1962). Hypoxemia in man induced by prolonged acceleration. *Acta-Physiol. Scand. 54*: 128-137.

Beck, B. G. & Tripp, L. D. (April 1989). *Use of lower body negative pressure as a countermeasure to negative $+G_z$ acceleration (AAMRL-TR-89-025).* Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Berkhout, J., O'Donnell, R. D., & Leverett, S. (June 1973). *Changes in electroencephalogram spectra during repeated exposure to $+G_z$ acceleration (AMRL-TR-72-123)*. Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

Braunstein, M. L. & White, W. J. (1962). Effects of acceleration on brightness discrimination. *Journal of the Optical Society of America, 52*(8), 931-933.

Buhman, J. R. & Perry, C. E. (1994). Human and manikin head/neck response to $+G_z$ acceleration when encumbered by helmets of various weights. *Aviation, Space, and Environmental Medicine, 65*: 1086-90.

Bulbulian, R., Crisman, R. P, Thomas, M. C., & Meyer, L. G. (1994). The effects of strength training and centrifuge exposure on $+G_z$ tolerance. *Aviation, Space, and Environmental Medicine, 65*; 1097-1104.

Burns, J. W., Werchan, P. M., Fanton, J. W., & Dollins, A. B. (1991). Performance recovery following $+G_z$-Induced Loss of Consciousness. *Aviation, Space, and Environmental Medicine*, 615-617.

Burns, J. W. & Whinnery, J. E. (1984). Significance of headrest geometry in $+G_z$ protective seats. *Aviation, Space, and Environmental Medicine, 55* (2): 122-127.

Burns, J. W. (1988). Prevention of loss of consciousness with positive pressure breathing and supenating seat. *Aviation, Space, and Environmental Medicine*, 20-22.

Burton, R. R. (1980). Human responses to repeated high G simulated aerial combat maneuvers. *Aviation, Space, and Environmental Medicine, 51* (W), 1185-1192.

Burton, R. R. (1986). A conceptual model for predicting pilot group G tolerance for tactical fighter aircraft. *Aviation, Space, and Environmental Medicine*, 733-744.

Burton, R. R. (1988). Anti-G suit inflation rate requirements. *Aviation, Space, and Environmental Medicine, 59*: 601-605.

Burton, R. R. (1988). G-Induced Loss of Consciousness: Definition, history, current status. *Aviation, Space, and Environmental Medicine*, 2-5.

Burton, R. R.; Hamalainen, O.; Kuronen, P.; Hanada, R.; Tachibana, S.; Henriksen, I.; Holewijn, M.; Klukowski, K.; Kwasucki, J.; Marke, J.; Mazurek, J; Talar, J.; Bukhtiyarov, I.V.; Khomenko, M.; Stupakov, G.; Vasilyev, A.; Diaz, C.; Garcia, S.; Del Valle Garrido Cano Herrera, M.; Tejada, F. R.; Harms-Ringdhal. K.; Linder, J.; Petren-Mallin Spangberg, C.; Alan, M.; Buyukcakir, C.; Sarykayalar, U.; Savasan, M.; Turbedar, E; Yilmaz, U.; & Oakley, C. (February 1999). Cervical spinal injury from repeated exposures to sustained acceleration. North Atlantic Treaty Organization Research and Technology Organization (RTO-TR-4 AC/323 (HFM) TP/9). Neuilly-sur-Seine Cedex, France: NATA RTO.

Burton, R. R.; Leverett, S. D.; & Michaelson, E. D. (1974). Man at high sustained $+G_z$ acceleration: a review. *Aerospace Medicine*, 1115-1136.

Burton, R. R. & Shaffstall, R. M. (1980). Human tolerance to aerial combat maneuvers. *Aviation, Space, and Environmental Medicine, 51* (7), 641648.

Cammarota, J. P. (November 1985). *A training method for the dynamic tracking of peripheral vision limits (NADC-86072-60).* Washington, DC: Naval Air Systems Command.

Cammarota, J. P. (1990). Evaluation of full-sortie closed-loop simulated aerial combat maneuvering on the human centrifuge. Proceedings of the IEEE 1990 National Aerospace and Electronics Conference, 838-842.

Canfield, A. A.; Comrey, A. L.; Wilson, R. C. (1949). A study of reaction time to light and sound a related to increased positive radial acceleration. *Journal of Aviation Medicine, 20* (6), 350-355.

Chambers, R. M. (1963). Operator performance in acceleration environments. In NM Burns, RM Chambers, and Hendler, Eds. Unusual environments and human behavior. London: The Free Press of Glencoe.

Chambers, R. M. & Hitchcock, L. (March 1963). *Effects of acceleration on pilot performance (NADC-MA-6219)*. Johnsville, PA: United States Naval Air Development Center.

Chelette, T. L.; Albery, W. B.; Esken, R. L.; & Tripp, L. D. (1998). Female exposure to high G: performance of simulated flight after 24 hours of sleep deprivation. *Aviation, Space, and Environmental Medicine, 69* (9), 862-868.

Chelette, T. L.; Albery, W. B.; McClosky, K. A.; & Goodyear, C. D. (April 1996). Task performance throughout prolonged high G exposure. Proceedings of the Neurological Limitation of Aircraft Operations: Human Performance Implications Conference. AGARD-CP-579, 3-1 to 3-4.

Cheung, B. & Hofer, K. (1999). Degradation of visual pursuit during sustained +3 G_z acceleration. *Aviation, Space, and Environmental Medicine, 70* (5), 451-458.

Convertino, V. A. & Reister, C. A. (2000). Effect of G-suit protection on carotid-cardiac baroreflex function. *Aviation, Space, and Environmental Medicine, 71*: 31-36.

Creer, B. Y.; Stewart, J. D.; & Douvillier, J. G. (1962). Influence of sustained accelerations on certain pilot-performance capabilities. *Aerospace Medicine*, 1086-1093.

Darrah, M. I. & Klein, E. A. (1987). Simulation of a highly dynamic G-time profile - a predictive algorithm for crewmember acceleration tolerance. Proceedings of the SAFE Association Annual Symposium; 15-20.

Darwood, J. J.; Repperger, D. W.; & Goodyear, C. D. (1991). Mass discrimination under G_z acceleration. *Aviation, Space, and Environmental Medicine, 62*: 319-324.

Drew, W. E. D. (2000). Spinal symptoms in aviators and their relationship to G-exposure and aircraft seating angle. *Aviation, Space, and Environmental Medicine, 71*: 22-30.

Edelberg, R.; Henry, J. P.; Maciolek, J. A.; Salzman, E. W.; & Zuidema, G. D. (1956). Comparison of human tolerance to accelerations of slow and rapid onset. *Aviation Medicine*, 482-489.

Firth, J. (1993, June). A plain pilot's guide to GILOG. *Sport Aerobatics*: 35-42.

Forster, E. M. & Whinnery, J. E. (1988). Recovery from G_z-induced loss of consciousness: psychophysiologic considerations. *Aviation, Space, and Environmental Medicine*, 517-522.

Forster, E. M.; Cammarota, J. P.; & Whinnery, J. E. (1994). G-LOC recovery with and without G-suit inflation. *Aviation, Space, and Environmental Medicine, 65*: 249-53.

Forster, E. M. & Cammarota, J. P. (1993). The effect of G-LOC on psychomotor performance and behavior. *Aviation, Space, and Environmental Medicine*, 132-138.

Forster, E. M. (1994). Heart rate response of aircrew during recovery from gradual onset rate +Gz exposures. *Aviation, Space, and Environmental Medicine*, 55-59.

Frazier, J. W.; Repperger, D. W.; & Popper, S. E. (1990). Time estimating ability during G_z. *Aviation, Space, and Environmental Medicine*, 449.

Frazier, J. W.; Repperger, D. W; Toth, T. N.; & Skowronski, V. D. (1982). Human tracking performance changes during combined + G_z and ? G_y stress. *Aviation, Space, and Environmental Medicine, 53*(s): 435-439.

Gaines, M. (1987, 28 March). G-LOC: taming the killer. *Flight International*, 27-30.

Gillingham, K. K. (1986). *G-tolerance standards for aircrew training and selection (SAM-TR-86-12)*. Brooks Air Force Base, TX: School of Aerospace Medicine.

Gillingham, K. K. (1987). G-tolerance standards for aircrew training and selection. *Aviation, Space, and Environmental Medicine*, 1024-1026.

Gillingham, K. K. (1988). High-G stress and orientational stress: physiologic effects of aerial maneuvering. *Aviation, Space, and Environmental Medicine*, A10-A20.

Gillingham, K. K. & Fosdick, J. P. (1988). High-G training for fighter aircrew. *Aviation, Space, and Environmental Medicine*, 12-19.

Gillingham, K. K. & Krutz, R. W. (1974). *Effects of the abnormal acceleratory environment of flight (SAM-TR-74-57)*. Brooks Air Force Base, TX: School of Aerospace Medicine.

Gillingham, K. K.; Makalous, D. L.; & Tays, M. A. (1982). G stress on A-10 pilots during JAWS II exercises. *Aviation, Space, and Environmental Medicine*, 336-341.

Gillingham, K. K.; Plentzas, S.; & Lewis, N. L. (July 1985). *G environments of F-4, F-5, F-15, and F-16 aircraft during F-15 tactics development and evaluation (USAFSAM-TR-85-51)*. Brooks Air Force Base, TX: United States Air Force School of Aerospace Medicine.

Gillingham, K. K.; Schade, C. M.; Jackson, W. G.; & Gilstrap, L. C. (1986). Women's G tolerance. *Aviation, Space, and Environmental Medicine*, 745-753.

Goodman, L. S.; Banks, R. D.; Grissett, J. D.; & Saunders, P. L. (2000). Heart rate and blood pressure responses to +G_z following varied-duration -G_z. *Aviation, Space, and Environmental Medicine, 71*: 137-141.

Goodman, L. S.; de Yang, L.; Kelso, B.; & Liu, P. (1995). Cardiovascular effects on varying G-suit pressure and coverage during +1 Gz positive pressure breathing. *Aviation, Space, and Environmental Medicine, 66*: 829-836.

Goodman, L. S.; Fraser, W.D.; Ackles, K. N.; Mohn, D.; & Pecaric, M. (1993). Effect of extending G-suit coverage on cardiovascular responses to positive pressure breathing. *Aviation, Space, and Environmental Medicine, 64*:1101-7.

Gray, G. W. & Paul, M. A. (April 1996). Seizure associated with G-LOC: A potentially fatal occurrence. *Proceedings of the Neurological Limitation of Aircraft Operations: Human Performance Implications Conference*. AGARD-CP-579, 2-1 to 2-4.

Green, N. D. C. (1997). Arm arterial occlusion cuffs as a means of alleviating high $+G_z$ – associated arm pain. *Aviation, Space, and Environmental Medicine, 68*: 715-721.

Griffin, M. J. (1975). Levels of whole-body vibration affecting human vision. *Aviation, Space, and Environmental Medicine, 46*, 1033-1040.

Guignard, J. C. & Irving, A. (1960). Effects of low-frequency vibration on Man. *Engineering Index, 51*.

Hamalainen, O.; Torvakka-Hamalainen, S. K.; & Kuronen, P. (1999). $+G_z$ associated stenosis of the cervical spinal canal in fighter pilots. *Aviation, Space, and Environmental Medicine, 70*: 330-334.

Hamalainen, O.; Vanharanta, H.; & Bloigu, R. (1994). $+G_z$-related neck pain: a follow-up study. *Aviation, Space, and Environmental Medicine, 65*: 16-8.

Hamalainen, O.; Vanharanta, H.; Hupli, M.; Karhu, M.; Kuronen, P.; & Kinnunen, H. (1996). Spinal shrinkage due to $+G_z$ forces. *Aviation, Space, and Environmental Medicine, 67* (7): 659-61.

Hamalainen, O.; Vanharanta, H.; & Kuusela, T. (1993). Degeneration of cervical intervertibral disks fighter pilots frequently exposed to high $+G_z$ forces. *Aviation, Space, and Environmental Medicine*, 692-696.

Harding, R. M. & Bomar, J. B. (1990). Positive pressure breathing for acceleration protection and its role in prevention of in-flight G-induced loss of consciousness. *Aviation, Space, and Environmental Medicine*, 845-849.

Hendriksen, I. J. M. & Holewijn, M. (1999). Degenerative changes of the spine of fighter pilots of the Royal Netherlands Air Force (RNLAF). *Aviation, Space, and Environmental Medicine, 70* (11), 1057-1063.

Hornick. R. J. (1973). Vibration. In JF Parker and VR West (Eds.) Bioastronautics data book. Washington, DC, National Aeronautics and Space Administration (USA Government Printing Office, Washington DC, Stock Number 3300-00474).

Hrebien, L. (1983). *Blood flow measurements under high-G conditions: early prediction of G tolerance.* Warminster, PA: Naval Air Development Center, Report No. MADC-83115-60.

Hull, D. H.; Wolthuis, R. A.; Gillingham, K. K.; McCracken, J. W.; & Triebwasser, J. H. (1977). *Effect of age on relaxed $+G_z$ tolerance of aircrewman*. AGARD Proceedings of the Specific Findings in Cardiology and Pulmonary Function with Special Emphasis on Assessment Criteria for Flying, B14-1 - B14-4.

Ioseliani, K. K. (1967). *The effect of vibration and noise on the mental faculty of man under time stress (NASA-TT-F-11-142)*. Washington DC: National Aeronautics and Space Administration.

Johanson, D. C.; Flick, V. P.; & Terry, D. M. (1986). An in-depth look at the incidence of in-flight loss of consciousness within the U. S. Naval Service: a final report (TP 6737). China Lake, CA: Naval Weapons Center.

Johanson, D. C. & Phenny, H. T. (1987). A new look at the loss of consciousness experience within the U. S. Naval forces. *Aviation, Space, and Environmental Medicine*, 6-8.

Kerr, W. K. & Russell, W. A. M. (April 1944). *Effects of positive acceleration in the centrifuge and in aircraft on functions of the central nervous system (C.2719).* National Research Council of Canada: Associate Committee on Aviation Medicine Research.

Kikukawa, A.; Miyamoto, Y.; & Akamatsu, T. (1966). $+G_z$ acceleration affects the trace minerals zinc, copper, and chromium in serum and urine of humans. *Aviation, Space, and Environmental Medicine, 67*;1166-1169.

Kikukawa, A.; Tachibana, S.; & Yagura, S. (1995). G-related musculoskeletal spine symptoms in Japan Air Self Defense Force F-15 pilots. *Aviation, Space, and Environmental Medicine, 65*; 269-272.

Krock, L. P.; Balldin, U. I.; Harms-Ringdahl, K.; singstad, C. P.; Linder, J.; & Siegborn, J. (1997). Influence of a reduced G-suit pressure schedule on G-duration tolerance using enhanced G-protection ensembles. *Aviation, Space, and Environmental Medicine, 68* (5), 403-409.

Lambert, E. H. & Wood, E. H. (July 1946). *The problem of blackout and unconsciousness in aviators.* The Medical Clinics of North America, Mayo Clinic Number, 833-844.

Leverett, S. D.; Burton, R. R.; Crossley, R. J.; Michaelson, E. D.; & Shubrooks, S. J. (December 1973). Physiological responses to high sustained $+G_z$ acceleration (SAM-TR-73-21). Brooks Air Force Base, TX: United States Air Force School of Aerospace Medicine.

Lewis, C. H. & Griffin, M. J. (1978). A review of the effects of vibration on visual acuity and continuous manual control. *Journal of Sound and Vibration, 56*, 415-457.

Lewis, C. H. & Griffin, M. J. (1979). The effect of character size on the legibility of a numeric display during vertical whole body vibration. *Journal of Sound and Vibration, 57*, 562-565.

Linde, L. & Balldin, U. (1998). Arm pain among Swedish fighter pilots during high +Gz flight and centrifuge exposures. *Aviation, Space, and Environmental Medicine, 69* (7), 639-642.

Ludwig, D. A.; Krock, L. P.; Doerr, D. A.; & Convertino, V. A. (1998). Mediating effect of onset rate on the relationship between +Gz and LBNP tolerance and cardiovascular reflexes. *Aviation, Space, and Environmental Medicine, 69* (7), 630-638.

Lyons, T. J.; Harding, R.; Freeman, J.; & Oakley, C. (1992). G-induced loss of consciousness accidents: USAF experience 1982-1990. *Aviation, Space, and Environmental Medicine*, 60-66.

Lyons, T. J.; Marlowe, B. L.; Michaud, V. J.; & McGowan, D. J. (1997). Assessment of the anti-G straining maneuver (AGSM) skill performance and reinforcement program. *Aviation, Space, and Environmental Medicine, 68*: 322-324.

McCloskey, K. A.; Tripp, L. D.; Chelette, T. L.; & Popper, S. E. (1992). Test and evaluation metrics for use in sustained acceleration research. *Human Factors, 34*(4): 409-428.

McKenzie, I. & Gillingham, K. (1993). Incidence of cardiac dysrhythmias occurring during centrifuge training. *Aviation, Space, and Environmental Medicine, 64*(8): 687-691.

McLeod, P.; Poulton, C.; duRoss, H.; & Lewis, W. (1980). The influence of ship motion on manual control skills. *Ergonomics, 23*, 623-634.

Martin, D.S.; D'Aunno, D.S.; Wood, M.L.; & South, D.A. (1999). Repetitive high G exposure is associated with increased occurrence of cardiac valvular regurgitation. *Aviation, Space, and Environmental Medicine, 70*(12), 1197-2000.

Meeker, L. J. (1991). Effects on G_z endurance/tolerance of reduced pressure schedules using the Advanced Technology Anti-G Suit (ATAGS). *Proceedings of the High Altitude and High Acceleration Protection for Military Aircrew Conference. AGARD-CP-516*, 15-1 to 15-4.

Michaud, V. J. & Lyons, T. J. (1998). The "push-pull effect" and G-induced loss of consciousness in the U.S. Air Force. *Aviation, Space, and Environmental Medicine, 69*(11), 1104-1106.

Michaud, V.J.; Lyons, T. J.; & Hansen, C. M. (1998). Frequency of the "push-pull effect" in U.S. Air Force Fighter Operations. *Aviation, Space, and Environmental Medicine, 69*(11), 1083-1086.

Moore, T. W.; Jaron, D.; Hrebien, Z.; & Bender, D. (1993). A mathematical model of G time-tolerance. *Aviation, Space, and Environmental Medicine, 64*: 947-951.

Moseley, M. J. & Griffin, M. J. (1986). Effects of display vibration and whole body vibration on visual performance. *Ergonomics 29*, 977-983.

Newman, D. G. (1997). +G_z induced neck injuries in Royal Australian Air Force fighter pilots. *Aviation, Space, and Environmental Medicine, 68*(6), 520-524.

Newman, D. G. (1997). Head positioning for high +G_z loads: an analysis of the techniques used by F/A-18 pilots. *Aviation, Space, and Environmental Medicine, 68*: 732-735.

Newman, D. G. & Callister, R. (1999). Analysis of the G_z environment during air combat maneuvering in the F/A-18 fighter aircraft. *Aviation, Space, and Environmental Medicine, 70*: 310-315.

Newman, D. G.; White, S. A.; & Callister, R. (1999). Patterns of physical conditioning in Royal Australian Air Force F/A-18 pilots and the implications for +G_z tolerance. *Aviation, Space, and Environmental Medicine, 70*: 739-744.

Newman, D. G.; White, S. W.; & Chevette, R. (1998). Evidence of baroflex adaptation to repetitive +G$_z$ in fighter pilots. *Aviation, Space, and Environmental Medicine, 69* (5), 446-451.

Newsom, W.A.; Leverett, S. D.; & Kirkland, V. E. (1968). Retinal circulation in man during centrifugal acceleration. *Transactions of the American Academy of Ophthalmology and Otolaryngology*, 39-49.

Nunnelly, S. A.; French, J.; Vanderbeek, R. D.; & Stranges, S. F. (1995). Thermal study of anti-G ensembles aboard F-16 aircraft in hot weather. *Aviation, Space, and Environmental Medicine, 66*: 309-312.

Orrick, W. P.; York, P. E.; & Cohen, M. M. (1976). *Human performance under acceleration: Actuation of ejection seat lower firing control (NADC-75268-40)*. Washington, DC: Naval Air Systems.

Pancratz, D. J.; Bomar, J. B. Jr.; Raddin, J. H. Jr. (1994). A new source for vestibular illusions in high agility aircraft. *Aviation, Space, and Environmental Medicine, 65*: 1130-3.

Park, J. K.; Seul, K. H.; Park, B. O.; Kim, S. H.; & Cho, K. W. (1994). Effects of positive acceleration on atrial matriuretic-peptide in humans. *Aviation, Space, and Environmental Medicine, 65*; 51-4.

Parkhurst, M. J.; Leverrett, S. D.; & Shubrooks, S. J. (1972). Human tolerance to high, sustained +G$_z$ acceleration. *Aerospace Medicine*, 708-712.

Paul, M. A. (1996). Instrument flying performance after G-induced loss of consciousness. *Aviation, Space, and Environmental Medicine, 67* (11); 1028-33.

Piranian, A. G. (August 1974). *The effects of sustained acceleration, airframe buffet, and aircraft flying qualities on tracking performance (AIAA Paper Number 74-793*. American Institute of Aeronautics and Astronautics Mechanics and Control of Flight Conference.

Pluta, J. C. (1984). LOC survey. *Flight Safety*. 25-28.

Popper, S. E. & Tripp, L. D. (1991). Unequal narrowing of the visual field in a +G$_z$ environment. *Aviation, Space, and Environmental Medicine*, 986-988.

Prior, A. R. J. (1987). R. A. F. experience of G-induced loss of consciousness. *Proceedings of the Royal Aeronautical Society High G and High G Protection Symposium*; 16-25.

Prior, A. R. G. (1991). The optimisation of a positive pressure breathing system for enhanced G protection. *Proceedings of the High Altitude and High Acceleration Protection for Military Aircrew Conference. AGARD-CP-516*, 14-1 to 14-11.

Rogers, D. B.; Ashare, A. B.; Smiles, K. A.; Frazier, J. W.; Skowronski, V. D.; & Holden, F. M. (1973). *Effect of modified seat angle on air-to-air weapon system performance under high acceleration (AMRL-TR-73-5)*. Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

Rositano, S. A. (February 1980). *Objective measurement of human tolerance to +G$_z$ acceleration stress*. NASA Technical Memorandum 81166.

Sadoff, M. & Dolkas, C. B. (1967). Acceleration stress effects on pilot performance and dynamic response. *IEEE Transaction on Human Factors in Electronics; HFE-8* (2): 103112.

Schafer, L. E. & Bagian, J. P. (1993). Overhead and forward reach capability during exposure to +1 to +6 G_x loads. *Aviation, Space, and Environmental Medicine*, *64*: 979-84.

Schmitz, M. A.; Simons, A. K.; & Boettcher, C. A. (1960). *The effect of low frequency, high amplitude, whole body vibration on human performance.* Report 130, Contract SA-49-007-MD-797. Milwaukee, WI: Bostraom Research Laboratories, (DTIC-AD: 241792).

Shamiss, A.; Meisel, S.; & Rosenthal, T. (1993). Acute hypertensive response to +G_z acceleration in mildly hypertensive pilots. *Aviation, Space, and Environmental Medicine*, *64*: 751-4.

Shoenberger, R. W. & Harris, C. S. (1981). *Cognitive performance during long-duration vibration (AMRL-TR-81-57).* Wright-Patterson Air Force Base, OH: USAF: Aerospace Medical Research Laboratory (DTIC-AD: A118612).

Shoenberger, R. W. & Wilburn, D. L. (1972). *Tracking performance during whole-body vibration with side-mounted and center-mounted control sticks (AMRL-TR-72-120).* Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory (DTIC-AD: 761798).

Stoll, A. M. (1956). Human tolerance to positive G as determined by the physiological end points. *Aviation Medicine*, 356-367.

Tachibana, S.; Akamatsu, T.; Nakamura, A.; & Yagura, S. (1994). Serious arrhythmias coinciding with alteration of consciousness in aircrew during +G_z stress. *Aviation, Space, and Environmental Medicine*, *65*: 60-66.

Tong, A.; Balldin, U. I.; Hill, R. C.; & Dooley, J. W. (1998). Improved anti-G protection boosts sortie generation ability. *Aviation, Space, and Environmental Medicine*, *69* (8), 117-120.

Tong, A.; Hill, R. C.; Tripp, L.; & Webb, J. T. (1994). The effect of head and body position on +G_z acceleration tolerance. *Aviation, Space, and Environmental Medicine*, *65* (5, Supplement): A90-4.

Travis, T. W. & Morgan, T. R. (1994). U.S. Air Force positive-pressure breathing anti-G system (PBG): subjective health effects and acceptance by pilots. *Aviation, Space, and Environmental Medicine*, *65* (5, Supplement): A75-9.

van Patten, R. E. (1983). *Investigation of the effects of G_y and G_z on AFTI/F-16 control inputs, restraints, and tracking performance* (AFAMRL-TR-83-080). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

von Gierke, H. E.; McCloskey, K.; & Albery, W. B. (1991). Military performance in sustained acceleration and vibration environments. Effects of acceleration and vibration. In R Gal and AD Mangelsdorff, Handbook of military performance, New York: Wiley and Sons, 335-364.

Watkins, S. M.; Welch, L.; Whitley, P.; & Forster, E. (1998). The design of arm pressure covers to alleviate pain in high G maneuvers. *Aviation, Space, and Environmental Medicine, 69* (5), 461-4667.

Webb, J. T.; Oakley, C. J.; & Meeker, L. J. (1991). Unpredictability of fighter pilot G tolerance using anthropometric and physiologic variables. *Aviation, Space, and Environmental Medicine*, 128-35.

Webb, P. (1964). *Bioastronautics data book* (NASA SP-3006). Washington, D C.: National Aeronautics and Space Administration.

Whinnery, C. C. M. & Whinnery, J. E. (1990). The effect of $+G_z$ offset rate on recovery from acceleration-induced loss of consciousness. *Aviation, Space, and Environmental Medicine*, 929-934.

Whinnery, J. E. (1986). *F-16 control stick response during $+G_z$-induced loss of consciousness* (SAMTR-86-26). Brooks Air Force Base, TX: School of Aerospace Medicine.

Whinnery, J. E. (1988). Converging research on $+G_z$-induced loss of consciousness. *Aviation, Space, and Environmental Medicine*, 9-11.

Whinnery, J. E. (1989). Observations on the neurophysiologic theory of acceleration $(+G_z)$ induced loss of consciousness. *Aviation, Space, and Environmental Medicine, 60*, 589-593.

Whinnery, J. E. (1990). Recognizing $+G_z$-induced loss of consciousness and subject recovery from unconsciousness on a human centrifuge. *Aviation, Space, and Environmental Medicine*, 406-411.

Whinnery, J. E. (1991). Comments on asymmetric visual loss associated with G_z-stress. *Aviation, Space, and Environmental Medicine, 62*: 575-576.

Whinnery, J. E. & Burton, R. R. (1987) $+G_z$-induced loss of consciousness: a case for training exposure to unconsciousness. *Aviation, Space, and Environmental Medicine*, 468-472.

Whinnery, J. E.; Burton, R. R.; Boll, P.A.; & Eddy, D. R. (1987). Characterization of the resulting incapacitation following unexpected $+G_z$-induced loss of consciousness. *Aviation, Space, and Environmental Medicine, 58* (7): 631-636.

Whinnery, J. E.; Fischer, J. R.; & Shapiro, N. L. (1989). Recovery to $+1 G_z$ and $+2 G_z$ following $+G_z$- induced loss of consciousness: operational considerations. *Aviation, Space, and Environmental Medicine*, 1090-1095.

Whinnery, J. E. & Gillingham, K. K. (1983). Medical standards for experimental human use in acceleration stress research. *Aviation, Space, and Environmental Medicine*, 241-245.

Whinnery, J. E.; Glaister, D. H.; & Burton, R. R. (1987). $+G_z$-induced loss of consciousness and aircraft recovery. *Aviation, Space, and Environmental Medicine*, 600-603.

Whinnery, J. E. & Jones, D. R. (1987). Recurrent $+G_z$-induced loss of consciousness. *Aviation, Space, and Environmental Medicine*, 943-947.

Whinnery, J. E. & Murray, D. G. (1990). *Enhancing tolerance to acceleration (+G$_z$) stress: the "Hook" maneuver* (NADC-90088-60). Warminster, PA: Naval Air Development Center.

Whinnery, J. E. & Shaffstall, R. M. (1979). Incapacitation time for +G$_z$-induced loss of consciousness. *Aviation, Space, and Environmental Medicine, 50*, 83-85.

White, W. J. (April 1960). *Variations in absolute visual thresholds during acceleration stress* (WADD Technical Report 60-34). Wright-Patterson Air Force Base, OH: Wright Air Development Division.

Whitley, P. E. (1997). Pilot performance of the anti-G straining maneuver: respiratory demands and breathing system effects. *Aviation, Space, and Environmental Medicine, 68*(4), 312-316.

Wiker, S. F.; Pepper, R. L.; & McCauley, M. E. *A vessel class comparison of physiological, affective state, and psychomotor performance changes in men at sea* (USCG-D-07-8). Washington, DC: United States Coast Guard (DTIC-AD: A098047).

Wilson, R. V. (1974). Display collimation under whole-body vibration. *Human Factors, 16*, 186-195.

Yilmaz, U.; Cetinguc, M.; & Akin, A. (1999). Visual symptoms and G-LOC in the operational environment and during centrifuge training of Turkish jet pilots. *Aviation, Space, and Environmental Medicine, 70*(7), 709-712.

Zuidema, G. D.; Cohen, S. I.; Silverman, A. J.; & Riley, M. B. (1956). Human tolerance to prolonged acceleration. *Aviation Medicine, 27*(6), 469-481.

Training Development Papers

Airline Pilot Training:
A University to Regional Airlines Bridge Training Model

Merrill R. Karp
Associate Professor/Associate Chair
Aeronautical Management Technology Department
Arizona State University
7442 East Tillman Avenue
Mesa, AZ 85212
Email: Ron.Karp@asu.edu

Abstract

Unfolding dynamics within United States' commercial aviation suggest that airlines and universities with airline-oriented flight training programs should form alliances to assure that the required quantity and quality of airline pilots are always available for regional airlines cockpits. Investment in implementing an integrated airline bridge training model has the potential to provide a substantial dividend in the quality of airline pilot production, to accelerate the supply of pilots into the regional airlines, and to provide clearer career pathways for university students in aviation programs. This model would include flight training focused on airline-type crew procedures and checklists, the use of specific airline flight training devices and motion-based simulators, and cooperative student candidate selection and employment interview agreements between regional airlines and universities.

Introduction

Commercial aviation within the United States is undergoing significant changes and challenges. Airline growth, retirements of Vietnam era military-trained commercial pilots, and a decrease in the resource of former military trained pilots to enter the airlines are factors to the developing pilot shortage in the U.S. airline industry. This shortage will eventually create an increased flow-through demand by the major airlines on the regional airlines for pilots, once the furloughed pilots from the recent economic downturn have been rehired by the major airlines or the regional airlines.

When the increased flow-through from the regional airlines to the major airlines does commence again, it will impact the regional airlines' training requirements and experience levels because qualified replacements may not be readily available. A pertinent commercial pilot supply issue to consider is that of the depth and quality of the aviation academic education and the structure and discipline of the flight training of those future airline pilots. Because of the increasing sophistication of modern aircraft and high technology equipment, this issue underscores a need to examine, and restructure, where necessary, the training options for potential airline pilots.

With this *examination and restructuring* in mind, the Aeronautical Management Technology Department at Arizona State University and Mesa Pilot Development, the training arm of Mesa Air Group, have developed and implemented an "Airline Bridge Training Model" after years of research and experimentation with the training model variables. While this Airline Bridge Training Model is <u>not</u> a "one-size-fits-all" model, regional airlines and universities and colleges with flight training programs can consider using it and customizing it into a model that fits their particular needs and resources.

This model details options for the cooperation, expectations, and agreements needed between universities and regional airlines to formalize a pipeline from the university academic and flight training environment to the regional airline cockpits. A major component of the airline bridge training model addresses aviation educational enhancements through the implementation of an integrated aviation learning model. The learning model incorporates elements of personal computer-based flight simulator programs, flight training devices, and full motion-based simulators, to span the gap between the academic classroom and the flight line. The model also incorporates the use of computer-based training programs and web-based training, as well as elements of the adult education paradigm, learning style theory, cooperative and collaborative learning techniques, to further enhance the knowledge transfer process. The importance placed on the use of PC-based flight simulator programs, flight training devices, and full motion-based simulators emphasizes providing immediate, hands-on application following each academic class, and is directed toward improving understanding and long-term retention, while increasing knowledge application across

a broad spectrum. Additionally, the airline bridge training model incorporates the early identification of those student pilots who have the greatest potential for success in commercial aviation by conducting extensive and on-going evaluation and assessment in cooperation with the airlines. An important element of the airline bridge training model is that subsequent pilot selection by the regional airlines must be contingent upon successful completion of both the university's aviation degree academic program and a flight training program which has a specific focus on airline-oriented checklists and procedures. Preferably, one cooperatively developed with the potential gaining regional airline.

The U.S. commercial airlines are starting to recover from the economic downturn over the past few years. To continue this recovery, it is important to review the background of where the airlines have been over the past few years, and then to consider the future projections as to where the airlines are going, as far as growth and hiring are concerned. The next step is to develop a unified airline industry and academia approach to plan for the future airline growth structure and personnel issues.

Background

Major/Regional Airlines' and Military Pilot Needs & Requirements

While the aviation industry has suffered a setback in the aftermath of the September 11, 2001, terrorist attack, a Federal Aviation Administration (FAA) 2003 Aerospace Forecast for FY 2003-2014 (Federal Aviation Administration, 2003) indicated that the U.S. and world aviation economies should start to recover by late 2003 and achieve moderate sustained growth through at least 2014. Although the final 2003 statistics are not available at the date of this article, the FAA did project that both the U.S. large carrier domestic and international passenger traffic are expected to achieve positive growth in 2003, with the international markets forecast to grow even faster than the domestic markets (4.7% versus 3.5 percent annually) over the 12-year projection period. The long-term impact of the September 11th attack on regional/commuter carriers (those air carriers whose majority of flights are operated in aircraft having 70 or less seats) has been generally more positive than negative. Regional/commuter passenger traffic is anticipated to grow at a faster rate than their larger domestic counterparts (5.6% versus 3.5% annually) over the 12-year FAA forecast period. Stronger growth for the regional/commuter traffic results from additional route transfers from their larger code-share partners and the increased purchasing/ leasing of new regional jets, as well as from the establishment of nontraditional point-to-point routes using the new regional jets. Over the past two years, regional/commuter airlines have reduced their fleet of piston and turboprop aircraft by 215 and replaced them with 462 regional jets capable of this more efficient routing, as well as increased passenger comfort. The number of regional jets in regional/commuter service is projected to grow from 976 in 2002 to 2,834 in 2014, an average increase of 9.3 percent annually (Federal Aviation Administration, 2003).

More commercial flights mean an increase in the requirement for pilots. In 2000, 19,027 new airline pilots were hired; this was a record number. In 2001, 12,766 new airline pilots were hired. In 2002, in spite of the slowdown in the aftermath of the September 11, 2001 terrorist attacks, 5845 new airline pilots were hired (Darby 2003a, Darby, 2002a; Darby, 2002b; Darby, 2001b). The downtrend now appears to have reversed, with an estimated 7,075 new airline pilot positions projected by the time the 2003 numbers are available. That prediction for these new pilots' positions for 2003 could be higher, depending on the domestic and international economies during the last quarter of calendar year 2003, and the two other main forces, pilot retirement and airline growth (Darby, 2003c). To meet pilot hiring requirements, the major airlines traditionally draw their pilots from the military and the regional and commuter airlines. As a result of the long-term increase in requirements for pilots in the major airlines, the regional airlines in the past have seen increases in their annual pilot flow rate to the major airlines ranging from 60 to 100 percent; this movement to the major airlines significantly impacts the regional airlines training requirements and experience levels (Ballenger, 2001).

Airline pilot growth rates are further complicated by the losses of the large number of Vietnam era military-trained commercial pilots who will reach mandatory retirement at age 60 within the next few years. Pilot retirements will continue to increase as more existing pilots reach age 60, from about 1,310 in 2003 to 2,064 in 2008, with 11,010 pilot retirements in the next five years (Darby, 2003c). By 2010, U.S. airlines will have to retire 50% of their pilots, varying from 32% to 60% of the airlines' current pilot population, depending on the specific airlines (Darby, 2001a; Darby, 2001b). To further exacerbate the commercial airline pilot shortage to meet the growing demand, because of low military pilot production over the past ten years, the resource of military trained pilots reaching the end of their initial commitment to the U.S. military has shrunk in the face of the projected increased demand for commercial pilots. The proportion of new hires with military experience into the commercial airlines has dropped from over 90 percent in 1990, to under 50 percent in only ten years (Taylor, Moore, & Roll, 2000). In 2002, civilian-trained pilots made up 68% of the 5,845 pilots hired by the airlines (Darby, 2003b). Military pilots have traditionally been in demand by the airlines because: (1) the quantity and quality of their training can be readily verified, (2) the variety and complexity of their flying experiences are well documented, (3) they have a traceable track record so that past problems cannot be easily hidden, and (4) flying proficiency and commitment can be tracked via career progression (Taylor, Moore, & Roll, 2000). In a study in 1999, the U.S. Air Force projected a shortfall of pilots of over 2000 in FY2002 and to remain over 2000 pilots through 2007, when the pilot resource would start a very slow, 15+ year return toward, but never reaching, USAF pilot requirements (Department of the Air Force, 1999). This situation has not improved, due to the War in the Gulf in 2003 and the resulting increased pilot requirements. Faced with this reduction in the former military pilot resource, the normal way for the major airlines to grow would be to increase the hiring from the regional airlines, where

the pilots have already proven themselves by their experience. However, with the regional airlines being a finite resource, action should be taken now to plan for the point when the regional airlines' resource could be reduced below requirements. Screening individuals in high school for the propensity to succeed, and then encouraging and facilitating those individuals to attend collegiate aviation programs with airline-focused academic and flight training, should be done now to prepare for future needs in the airlines.

Screening for Potential Success in Aviation

Evaluating and assessing pilots prior to their entering flight training leading to an airline career plays a critically important factor in selecting potentially successful pilots for high capital investment flight training programs. The pilot selection and testing process is considered a key to the success of military pilot training and includes tests for general cognitive abilities, personality, psychomotor skills, and physical fitness to eliminate individuals who were less likely to succeed (Karp, 2003).

Many airlines outside of the United States have been using pre-training screening for a number of years because of their unique pilot selection issues (Karp, 2003). In many countries' situations, the airlines must pay for all training, so the airline wants the highest return in competition rates for their very costly training investment. This was the case for many years with Lufthansa Airlines, who has been using comprehensive screening programs since the 1950's with tremendous success. Their screening programs have resulted in an exceptionally high pilot training completion rate of more than 90% (Dr. Karsten Severin, Director of Psychology, Lufthansa German Pilots School, personal communication, March 3, 1995). The German Aerospace Research Institute (DLR) has been responsible for the screening for pilots for Lufthansa Airlines for over 40 years. This screening has resulted in selection criteria such that less than 10% of the applicants who passed the screening fail to complete the flight training. The DLR contended that if the total profile of knowledge, ability, and personality is at or above their normative group in all areas, the individual has an extremely high probability of being a successful airline pilot (Dr. Klaus-Martin Goeters, Director, Aviation and Space Psychology Department, German Aerospace Research Institute, personal interview, Hamburg, Germany, April 2, 1996). Damos (1996) stated that the batteries used to select pilot candidates often predict training performance, rather than operational performance. A recent search of the literature on pilot selection revealed that test batteries often assess, for example, personality traits (Retzlaff, King, Callister, Orme, & Marsh, 2002; Rodgers, Covington, & Jensen, 1999), physiological factors (Clark & Riley, 2001; Klokker, Brock-Nannestad, Mikines, Johnsen, Garhoj, & Vesterhauge, 1999; Leino, Leppaeluoto, Ruokonen, & Kuronen, 1999), or aptitude (Carretta, 2000). While the popular use of these personality inventories suggests their utility, Damos (1996) stated that their predictive validity is low, and that further research should be conducted to identify new methods and improve those that are currently in use. Furthermore, measures to predict operational performance should be de-

veloped to assist in pilot selection of high school students. In addition to screening for those who have the potential to succeed in collegiate aviation programs, an integrated academic and flight training program is essential.

.

The Integrated Aviation Learning Model

Considering the literature review, an integrated aviation learning model, the Aviation Education Reinforcement Option, or AERO model (Figure 1), was developed to increase retention and enhance application of aviation education, with a focus on airline flight operations (Karp, 2000). This model has been implemented in the Aeronautical Management Technology Department of Arizona State University with highly positive results (Karp, McCurry, McHenry, & Harms, 2000).
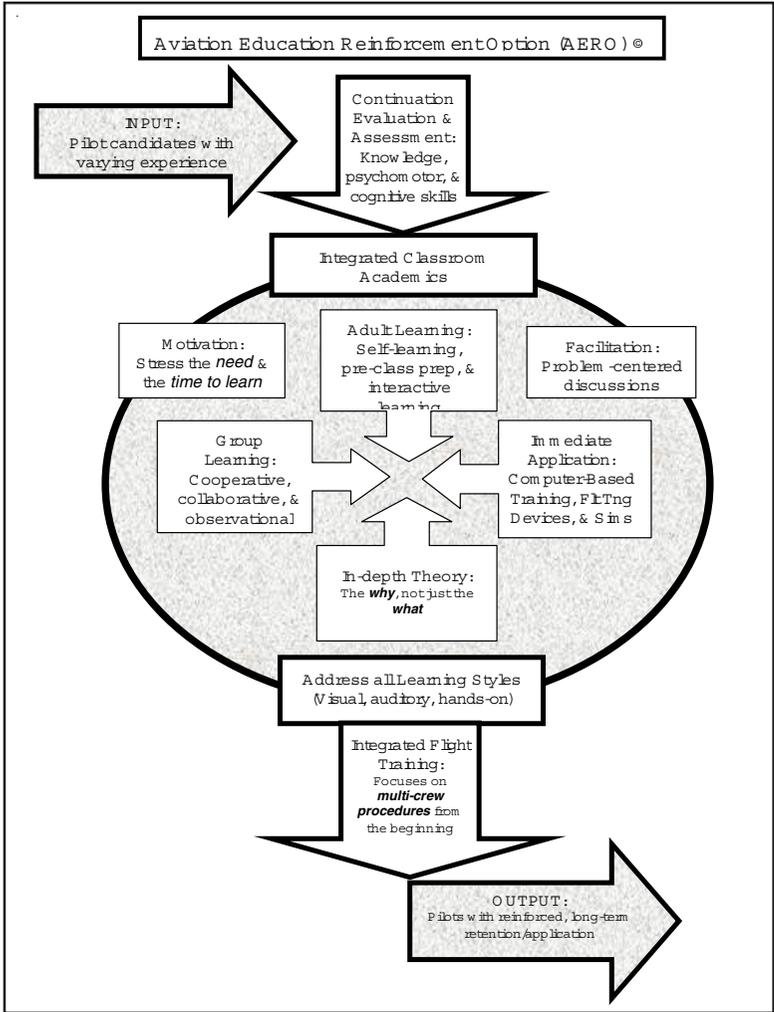

Figure 1.Aviation Education Reinforcement Option (AERO Model).

Integrated Learning Model Components

Inputs. While pilot candidates in a first officer training program can have varying levels of experience, university-age individuals, who have little or no flying experience, make excellent candidates because they have minimal "bad flying habits" or misconceptions.

Evaluation and Assessment Program. Potential first officer training program candidates should be tested and screened in advance for: psychomotor skills, temperament, information processing, and cognitive skills, as well as an FAA First-Class medical flight physical.

Integrated Aviation Classroom. Since university-age students are in a transition from adolescent learning to adult learning, beginning aviation students must be "focused" toward self-directed learning to attain their maximum potential. This includes motivating the learners by stressing the need to acquire the knowledge and to recognize that this is the time to learn it. While it is important that a lecture alone is effective when the learner has little or no knowledge of subject, facilitating the knowledge transfer is a more effective format to increase knowledge by engaging learners in an exchange of ideas in problem-centered discussions and tapping into their prior experiences (Brookfield, 1989; Cross, 1979; Zemke & Zemke, 1995).

Adult Education Principles. Goals for learning objectives and the methods for knowledge transfer and evaluation are important details for the educator to explain in order to assure a "buy-in" by the learners as to the "what" and "when" of the aviation learning process. Since adults cannot be "forced" to learn, it is important to emphasize that the pilots, themselves, must make that decision, and then "self-direct" the process (Brookfield, 1989). Observation of new freshmen entering into collegiate aviation programs indicates that this "self-direction" is not the model that most high school graduates use. It is pivotal that universities and colleges facilitate their students to move as quickly as possible into the adult learning model.

In-depth Theory. In order for pilots to apply recently acquired knowledge to new situations, they must have an in-depth understanding of systems and procedures. That is, a detailed comprehension of the why's, and not just the what's.

Immediate Application. Following each flight training classroom lesson, learners should go to a laboratory for immediate application of the lesson components to reinforce the knowledge transfer on personal computer-based flight simulator programs. Immediate, hands-on application of acquired knowledge is critical for adult learning and reinforcement to take place (Zemke & Zemke, 1995).

Group Learning. Group learning in small "praxis teams" (crews) is particularly applicable for aviation students. Group learning includes cooperative, col-

laborative, and observational learning. Cooperative learning takes place when the learner teams give presentations and fly simulator missions as assigned by the educator. In contrast, collaborative learning takes place when the educator makes an overall assignment to the group for presentations or flight simulator missions, and the group itself determines who will do what, and how (Bruffee, 1995; Matthew, Cooper, Davidson & Hawkes, 1995; Zemke & Zemke, 1995). In the collaborative learning laboratory, the teams should "fly" approaches or Line-Oriented Flight Training (LOFT) profiles, using "pilot-flying / pilot-not-flying" procedures early in their training to reinforce multi-crew concepts, as well as the airline oriented challenge-and-response type checklists and procedures. Collaborative learning has proven to be an especially reinforcing process for aviators. The observational learning element in group learning includes a non-flying team observing the team that is flying in the collaborative learning personal computer-based flight simulator laboratory. These observational teams then provide a post-simulator flight assessment. This group learning component provides direct peer feedback for the team who is flying, and objective observational learning for the non-flying team.

Learning Style Theory. Learning style theory, that is, the way people learn best, is of considerable importance in developing and delivering aviation academic programs. One model suggests that there are three recognized primary, or dominant, learning styles: (a) visual learners, who learn best by reading or looking at pictures; (b) auditory or aural learners, who learn best by listening; and (c) hands-on, tactile, or kinesthetic learners, who need to use their hands or whole body to learn (Filipczak, 1995). If knowledge transfer is to take place within the entire classroom population, all of these dominant learning styles should be addressed in the academic environment. Research of dominant learning styles (visual, auditory, or hands-on) underscores its importance to an integrated learning model. Over the past five years, 507 pilots (ranging from private pilots to F-16 pilots) were administered a written instrument to identify the respondents' dominant learning styles, as well as to explore potential enhancements and restructuring to aviation academic programs (Karp, Turney, Green, Sitler, Bishop, & Niemczyk, 2001; Karp, Condit, & Nullmeyer, 2000). The learning style assessment of the 507 pilots revealed that over 44% were hands-on learners, and almost 55% were either hands-on, or an equal combination of hands-on and/or visual or auditory learners. In contrast to the majority of the pilots being predominantly hands-on or an equal combination of hands-on and visual/auditory learners, the research indicated that most classroom instruction environments were auditory in nature with visual supplementation, but very little, if any, hands-on learning (Karp, Turney, Green, Sitler, Bishop, & Niemczyk, 2001). In developing educational programs, it is important to know how people learn the best, and why they succeed. Because of the depth and complexity of the subject matter, aviation academic instructors must present the course material in ways that satisfy the different needs and styles of the aviation learners. Likewise, each student must understand his or her dominant learning style and maintain more focused attention to the information when it is being presented in

a teaching style that is not easily compatible with their learning style. The application of learning style theory is a major component of the AERO model.

Integrated Flight Training. Integrated training focuses on multi-crew procedures from the beginning of flight training. By using airline-type procedures and checklists, pilots have minimal "procedural" transition issues when going to the airlines.

Output. The goal of the integrated aviation learning model is to produce a pilot who has long-term retention of the knowledge, and can successfully apply that knowledge to new situations without having previously encountered the new situation.

Airline Bridge Training Model

While an integrated aviation learning model, such as the AERO Model presented above, should enhance reinforced long-term retention and application and help accelerate qualified commercial pilot production, there is still a major gap in the process to move those new pilots into the regional and major airlines at a faster pace.

In a university flight program environment, the traditional pathway for a new pilot to enter eligibility to apply for the airlines is to build flight hours after graduation. This "flying hour-based experience" criterion has been the standard in the airlines for many years as the vehicle for the U.S. civil-trained future airline pilot to "grow" experience and "prove" his or her "motivation" to be in this demanding and challenging industry. This *flight time-based* track usually is accomplished first as a certified flight instructor (CFI) and then possibly as an air taxi pilot or a commuter pilot. There also are a number of other avenues to build flight hours, such as personally funded flying, crop dusting, banner towing, and traffic watch. Once pilots acquire enough flight time, they can apply to the regional airlines for employment. It is important to note though, that during the typical two to five years it takes to build the required hours to apply for the regional airlines, the pilot is <u>not</u> practicing or enhancing the airline focused multi-crew skills he or she had acquired in their university's flight training program.

In addition, when a regional airline does accept the pilot, the new hire has to be basically *retrained* in airline procedures and multi-crew relationships—making the pipeline to the regional airline cockpit even longer. In the regional airlines, they normally build additional flight time as quickly as possible to qualify to apply to the major airlines. With the airline pilot shortage now just on the horizon something must be done within the selection and hiring process to accelerate the flow of qualified pilots into the regional airlines (McCurry, Karp, Hayes, & Moman, 2002a).

Constructing a *training and employment bridge* from the university with an airline focused flight program, to the regional airlines would be a major step in spanning this gap. This bridge would be based on *proficiency* in flying skills and

airline procedures, in contrast to solely basing employment opportunity on *flight time* as a prerequisite for application. While a move in this direction would be a large step by many of the U.S. airlines, there are already some regional airlines addressing the issue of *proficiency* in lieu of *flight time*. For example, Mesa Air Group, through Mesa Pilot Development, has been employing for more than a decade, this airline-focused training, for individuals with no prior flight experience, for direct hiring into Mesa Air Group as First Officers. What is needed on a more national level, to bridge the commercial airline pilot supply and demand gap, is to establish formal relationships for training and employment bridges between major airlines, regional airlines, and universities, to link all of the components together (McCurry, Karp, Hayes, & Moman, 2002b).

For the *major airlines* to acquire pilots with the right skills and training, but with fewer flight hours, they could consider forming training and employment relationships with regional airlines to help establish training criteria and participate in a pre-training selection and screening process. This will help assure the quality of the pilots coming to them from their affiliated regional airlines. The major airlines could, in turn, offer the regional airlines' pilots the opportunity for employment processing after a *specified time-period* in the affiliated regional airlines. This will assure the major airlines of a pipeline of experienced pilots at known intervals and will allow the regional airlines to keep their experienced pilots for a predicable timeframe, helping the regional airlines to predict their training requirements and to recruit the correct number of pilots at the right time. Likewise, the *regional airlines* should form training and employment relationships with universities to establish the same training criteria and evaluation and assessment process. Similar to the major airlines' relationships with the regional airline pilots, the regional airlines could offer the opportunity for employment processing for the new pilots when they graduate from the university with which they have established a "First-Officer training program" pipeline.

The actions required to support this initiative require a *formal agreement* to assure that the needs and expectations of all of the participants (from the university students, to the regional airlines, to the major airlines) are met. If the investments are to be made by all parties concerned, they must be made with certain assurances.

The format to *Establish an Airline Bridge Training Model* (Figure 2) details some components of a training and employment relationship that could potentially accelerate the production of airline quality pilots from a university academic flight program toward eligibility for entry into a specific regional airline's first officer training program. The model would be based upon a four-year university aviation curriculum; a highly structured airline flight training program; continuous screening, evaluation, and assessment; and a hiring interview by the regional airline after successful completion of the bridge training program. In a parallel effort, the regional airline should establish similar agreements with major airlines to provide the quality and quantity of first officers that they need. The regional airlines would benefit by hiring a pilot who they helped select and who

was trained using procedures that they helped develop. Additionally the regional airlines could have the services of those pilots for a set time-period, as established in agreements with the major airlines with which they have aligned themselves. The major airlines gain by having an impact on the quality of training and experience received in the regional airlines and being able to forecast the available pilot pool from that regional.

At Arizona State University, this airline bridge training model has expanded on the success of the Mesa Air Pilot Development and San Juan College pro-
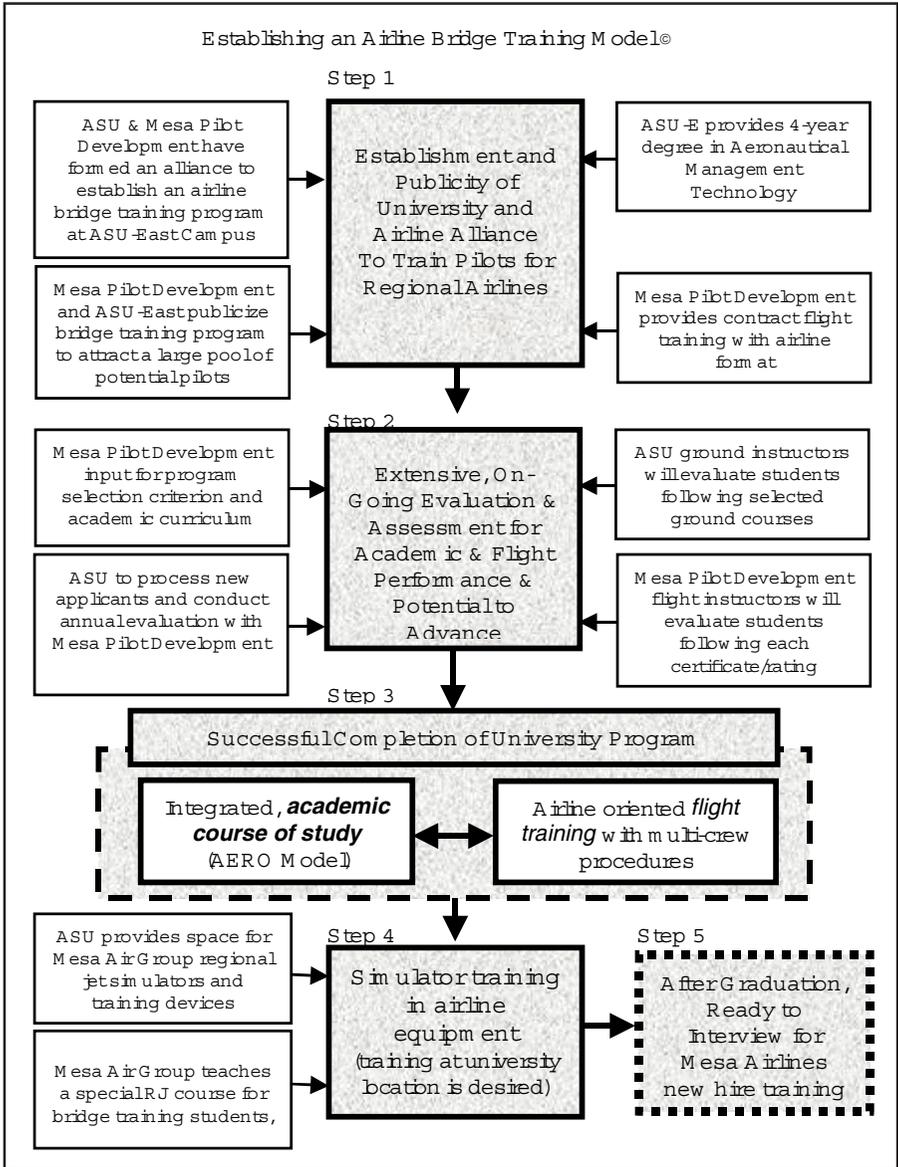

Figure 2. Establishing an Airline Bridge Training Model.

gram in Farmington, New Mexico. The Farmington flight training program has been in existence since 1990 and has been highly successful in producing pilots for direct hire into Mesa Air Group. For the past one and one-half years, Mesa Air Group has been placing graduates of the program immediately into the right seat of regional jets, with a 95% new hire pass rate. Overall, these graduates represent less than 1% of failures during all airline new hire flight training. Mesa Air Group feels that this success rate is due to the structured environment of their program and continuous student assessment. Mesa Air Group anticipates the same long-term results with the new airline bridge training program with ASU (Mickey Moman, Mesa Air Group Vice President for Training, Mesa, AZ, personal communication, August 8, 2001).

The first step in the Airline Bridge Training Model would be the establishment of the university and regional airline program. Because the competition for qualified pilots is getting increasingly more intense, these bridge training relationships must be highly publicized. In this area, because of limitations in most universities' budgets, the regional airlines will probably have to make most of the initial investment in advertising the newly formed training and employment program with the university.

Since Mesa Pilot Development was competitively selected as the flight contractor for the ASU flight program, it is in the contract that Mesa Pilot Development publicize and advertise the ASU program in connection with their other highly successful Mesa Pilot Development two-year program with San Juan College in Farmington, New Mexico. Arizona State University will recruit students for the four-year flight and academic program, leading to a Bachelor of Science degree in Aeronautical Management Technology, through the University's academic market and advertisement programs.

A major component of the airline bridge training model should be to seek out, attract, and recruit all of those individuals who have been evaluated and have demonstrated the potential to be successful airline pilots. In line with this effort, additional action must be taken to identify the scholarship funds and loans necessary to support the flight training and academic education for those who have the potential to be commercial airline pilots, but not necessarily the funds.

Airline Bridge Training Model Concepts

The academic and the flight training concept should based on airline procedures, including "challenge and response" and "callout" checklists, similar to those used by the airline partner, in the case of the ASU/Mesa Pilot Development Bridge program, the students use the same format and procedures as Mesa Air Group pilots flying the line in a regional jet airliner.

Students should fly in "team observation flights," with an instructor pilot on board, whenever possible to provide twice the flight situational exposure than single-pilot/instructor training. Students will switch seats after the first flight and the second student will fly while the second student observes. In the advanced

stages of training, in the regional jet flight training devices (FTD) and motion-based regional jet flight simulators, the second student crewmember will make all the "pilot-not-flying" radio calls and checklist callouts, just as they would when flying the line.

All academic ground courses go in-depth into the theoretical "why's" of aviation, not just the "what's." Extensive use of immediate hands-on application using PC-based flight simulator programs is used in the academic courses. Instrument pilot ground school, commercial pilot ground school, the advanced air navigation course, the airline instrument procedures course, and the certified flight instructor—instrument course all have flight simulator laboratories that extensively use PC-based Aviation Training Devices, with flight simulator equipment and programs, to reinforce the classroom instruction and to provide the student with interactive Air Traffic Control radio communications. Additionally, these classes use a computer based-training and video laboratory to emphasize airline systems, including glass cockpits, flight management systems, and global positioning systems.

Evaluation and Assessment

An essential component of any First Officer training program, is an intensive and on-going evaluation and assessment effort (Figure 3), which will be conducted jointly by the university and the regional airline. In such a bridge training program, the regional airlines must participate with the university in determining the selection criteria and training standards.

In the case of ASU/Mesa Air Group airline bridge training program, the selection program will take place over a four-year period, with certain specific annual milestones to be evaluated. For example, when students are accepted into the ASU flight program, they must be first accepted into Arizona State University, which has established minimum grade point averages from high school, ranking in high school class, SAT scores, etc. Once accepted into ASU, the student selects a degree program in Aeronautical Management Technology, with a "Professional Pilot" concentration, and registers in a number of set courses, e.g., private pilot ground school, meteorology, and the flight safety course that is associated with the private pilot certificate. The students also should be able to attain at least a first-class FAA flight physical. Senior members of the faculty from the ASU Department of Aeronautical Management Technology and Mesa Air Group/Mesa Pilot Development will conduct each of the phases of the four-year Airline Bridge Training Model Evaluation and Assessment, which also will include class attendance, professional attitude and conduct, and the student's driving record.

In the first semester private pilot flight safety course, if not given before the first semester starts, students will be given a series of assessment tests by Mesa Pilot Development and the AMT Department to determine the students' "potential for success." This initial assessment (Phase I), which will include a range of cognitive ability, temperament, and knowledge tests, will not be the
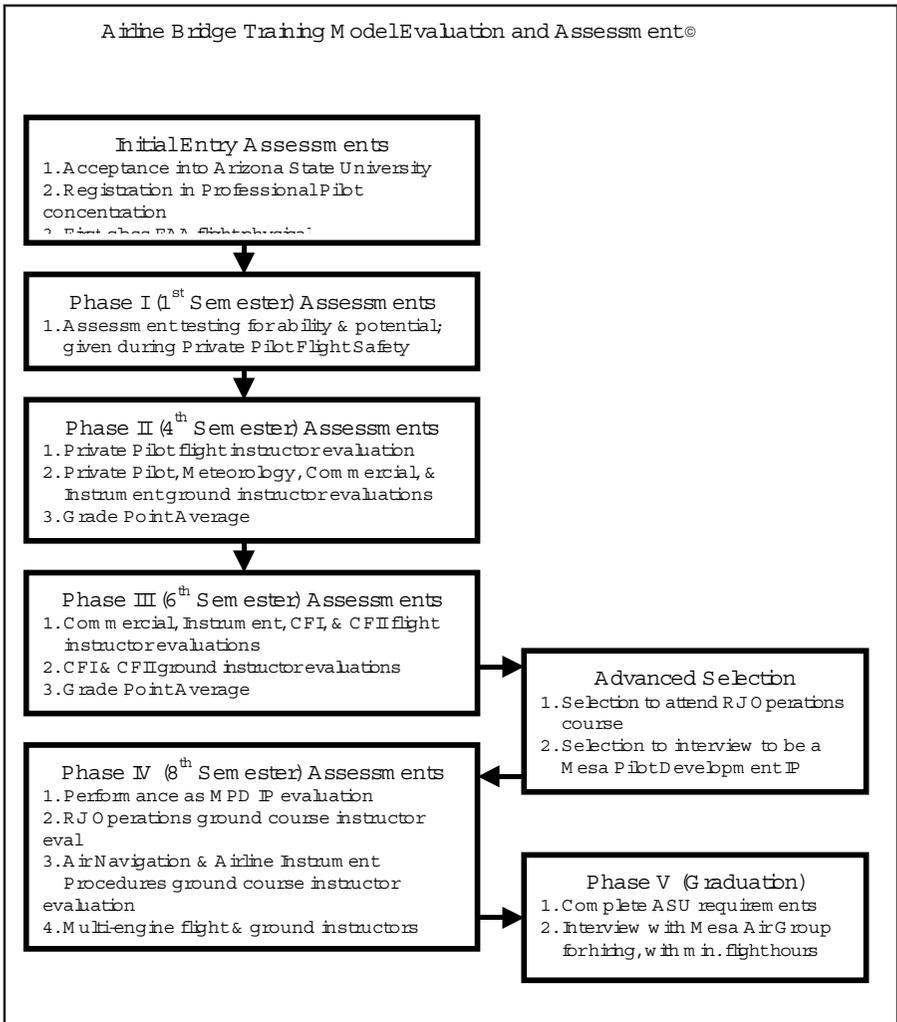
Airline Bridge Training Model Evaluation and Assessment©

Initial Entry Assessments
1. Acceptance into Arizona State University
2. Registration in Professional Pilot concentration
3. First class FAA flight physical

Phase I (1ˢᵗ Semester) Assessments
1. Assessment testing for ability & potential; given during Private Pilot Flight Safety

Phase II (4ᵗʰ Semester) Assessments
1. Private Pilot flight instructor evaluation
2. Private Pilot, Meteorology, Commercial, & Instrument ground instructor evaluations
3. Grade Point Average

Phase III (6ᵗʰ Semester) Assessments
1. Commercial, Instrument, CFI, & CFII flight instructor evaluations
2. CFI & CFII ground instructor evaluations
3. Grade Point Average

Advanced Selection
1. Selection to attend RJ Operations course
2. Selection to interview to be a Mesa Pilot Development IP

Phase IV (8ᵗʰ Semester) Assessments
1. Performance as MPD IP evaluation
2. RJ Operations ground course instructor eval
3. Air Navigation & Airline Instrument Procedures ground course instructor evaluation
4. Multi-engine flight & ground instructors

Phase V (Graduation)
1. Complete ASU requirements
2. Interview with Mesa Air Group for hiring, with min. flight hours

Figure 3. Airline Bridge Training Model Evaluation and Assessment.

single deciding factor on entry into the airline bridge training program; it is, however, a very important assessment. At the beginning of the fourth semester (second year) and the sixth semester (third year), all the students in each cohort "class" will be evaluated by a review board of ASU flight faculty and Mesa Pilot Development instructors. These evaluations are part of the on-going process to help the students to remain on track for selection for continued participation in the airline bridge training program.

In their fourth semester (second year), the students will be assessed (Phase II) using evaluations from their private pilot flight instructor and evaluations from their private pilot, meteorology, commercial, and instrument ground instructors, as well as their grade point average.

During their sixth semester (third year), the students will be assessed (Phase III) based on the evaluations by their commercial certificate, instrument rating, certified flight instructor (CFI) rating, and certified flight instructor-instrument (CFII) rating flight instructors. Additionally, they will be assessed by their CFI and CFII ground instructors and by their grade point average. At the conclusion of the sixth semester assessment, the selection will be made to determine if the students will attend the Regional Jet Operations course in their senior year. If a student is selected for the Regional Jet Operations Course, it will be a substitute for the Boeing 737 Airline Aircraft Systems course that is a flight program core curriculum requirement. Also, after the sixth semester evaluation, students who were selected for the Regional Jet Operations Course will have an opportunity to interview with Mesa Pilot Development to be a flight instructor.

Note: The Regional Jet Operations course will be modeled after a typical RJ new hire indoctrination course, but geared to the students' level of knowledge and experience, and will include some limited full-motion based RJ simulator time with the instructor. Students enrolled in the Regional Jet Operations course also will have access to Mesa Air Group's regional jet Pilot Training Modules (PTM), as well as RJ training manuals, videos, and computer based training.

The Regional Jet Operations Course will be focused on teaching the students about RJ equipment, systems, and operations, as well as presenting a further opportunity for Mesa Air Group to assess the ASU students as potential future employees. Attendance in this course will be allowed only if the students have received recommendations to continue in the bridge training program at the beginning of the sixth semester (third year), assuming that they have remained in their original "class" cohort group.

Note: Students not selected for the RJ Operation Course or to interview with Mesa Pilot Development for a flight instructor position are encouraged to instruct with other flight providers and continue their preparation to apply for regional airlines after graduation. The ASU Aeronautical Management Technology Department will work with any regional airlines that desire to establish internships or interview these extremely well qualified pilots.

At the end of the eighth semester (fourth year), each student will be assessed (Phase IV) by the Mesa Pilot Development chief pilot as to the student's performance as an instructor. This assessment will be made based on input by the Regional Jet Operations course ground instructor, by the ground instructors for air navigation and airline instrument procedures courses, by the multi-engine ground and flight instructors, and by their grade point averages.

As the pilot approaches graduation and successful completion of the university's bridge training program, the pilot is ready for advanced academics and simulator training in regional airline equipment. When possible, this should be accomplished at the university location as part of the final employment selection process.

Regional Jet Simulator Training

The regional airline should make the investment in placing a simulator at the university, which could serve also as the location for the regional airlines' periodic initial/recurrent simulator training. To make such an expensive, collaborative venture successful, the university could provide the academic classrooms and administrative space, as well the simulator space. Additionally, the airline simulator located at a university would be an extremely valuable recruiting and retention asset for both the university and the regional airlines' first officer training program. Also, an airline simulator located at a university would permit the regional airlines to establish indoctrination classes for their line pilots, for FAA credit, which could be taught by approved and certified university faculty, and flown in the on-site simulator.

In the case of the Arizona State University/Mesa Air Group airline bridge training program, ASU is providing Mesa Air Group with space for regional jet simulators, classrooms, and offices; accordingly, Mesa Air Group will arrange for regional jet simulators to be positioned at ASU for their pilots attending new hire, upgrade, recurrent simulator training. In return for this space, Mesa Air Group not only will teach the ASU Regional Jet Operations course, which includes time in the RJ motion-based level D simulator, but also will enroll their company pilots attending level-D simulator training at ASU as Arizona Statue University continuing education students. In February 2002, Mesa Air Group and CAE, the simulator provider, completed installation of CRJ and an ERJ level-D simulator in the Arizona State University simulator building, setting into motion a critical step in the Airline Bridge Training Model.

Selection for Hiring Interview

After the eight-semester evaluation and assessment process (in the fourth year), and successful completion of the ASU/Mesa Air Group Airline Bridge Training Program, the new graduates are eligible to submit an application to Mesa Air Group for a hiring interview. The specific assessment elements for review will be developed by Mesa Air Group and each incoming first semester freshman will be briefed by Mesa Air Group as to the depth of the assessment and evaluation process before they ever begin the airline bridge training program.

After successfully completing the advanced academics and simulator training and graduating from the university, the new pilot is ready to join the regional airlines, enter new-hire flight training, and then quickly start flying the line. These new pilots should be able to move directly into the right seat, armed with a working knowledge of current airline procedures, as well as the individual regional airlines' flight procedures and equipment specifics.

Arizona State University/Mesa Pilot Development
Airline Bridge Training Model

The complete *Airline Bridge Training Model* (Figure 4) uses the relationships

between Arizona State University and Mesa Air Group as an example for other universities, colleges, and regional airlines to customize for their specific needs and capabilities. ASU and Mesa Pilot Development, in cooperation Mesa Air Group, are conducting on-going research to validate the "predictability for success" of the initial testing and skills assessment of the new freshman. This eight-year longitudinal study assesses academic and flight performance during their undergraduate program and then considers their progress as a First Officer in Mesa Air Group until they become a Captain. The results of this research will be made available to academia and the airline industry as further conclusions are reached.
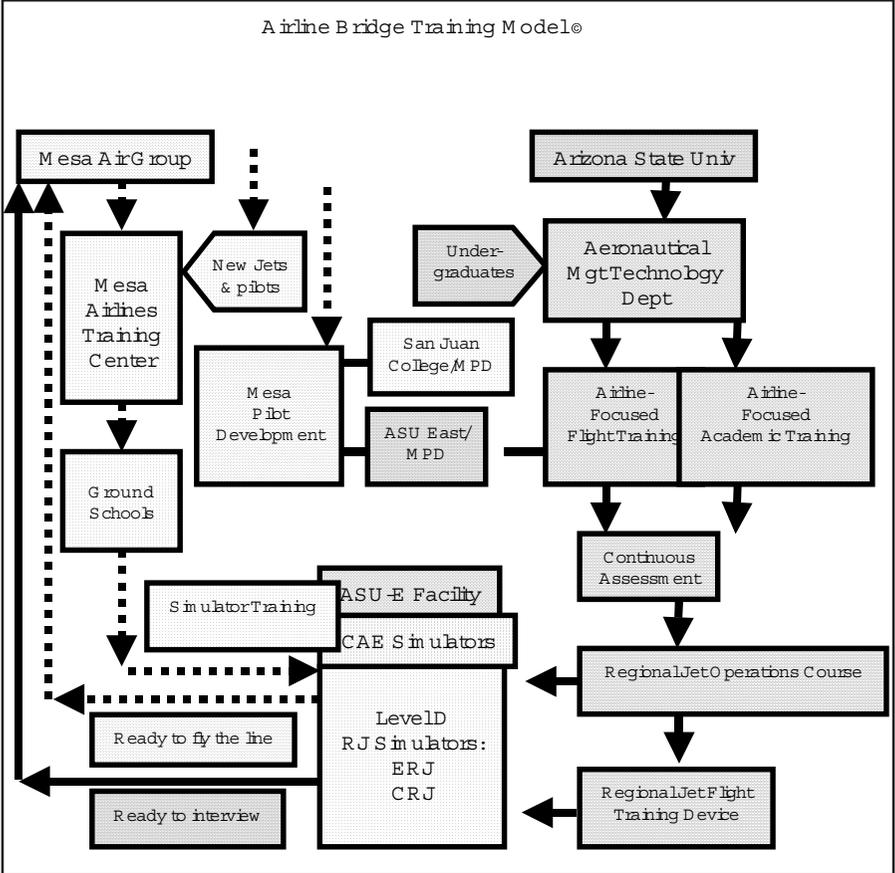
Figure 4. Arizona State University/Mesa Pilot Development Airline Bridge Training Model.

Summary

The first seven students graduating from the Arizona State University/ Mesa Pilot Development Airline Bridge Training Program submitted their applications in 2003 to Mesa Air Group and were subsequently hired as First Officers and assigned directly into regional jets. All of these new hire pilots have successfully completed training and are flying the line. The next class of graduates

has just started the application and interview process. Mesa Air Group feels that although the 100% success rate for the first group of graduates is a very good indicator of the effectiveness of the program, the sample size of seven is still too small to make any predictions that future graduates will have the same 100% hiring and training completion success. However, Mesa Air Group is very encouraged with the quality of this first airline bridge training program graduating class and feels that the jointly designed university/airline bridge training program is just what Mesa Air Group needs to help acquire new pilots who have had an airline-focused university education and flight training using specific Mesa Air Group checklists, procedures, and disciplines to be part of Mesa Air Group's rapidly growing regional jet fleet (Mr. Pete Hayes, President of Mesa Pilot Development, Mesa, AZ, personal communication, January 6, 2004).

Recommendations

1. Aviation education and training institutions should adopt an integrated aviation learning model, such as the AERO Model in Figure 1, which uses the adult education paradigm and cooperative and collaborative learning techniques, in concert with PC-based flight simulator programs and flight training devices for immediate classroom hands-on application of airline multi-crew cockpit procedures.

2. Formal relationships should be established between universities and regional airlines for participation in the pre-training selection process, training program development, and early identification for employment, as detailed in the Establishing Airline Bridge Training Model in Figures 2 and the full Airline Bridge Training Model in Figure 4.

3. A national evaluation and testing pre-screening process should be developed, validated, and implemented to assess interested high school students to see if they have the potential to be airline pilots. All individuals whose testing and evaluation indicate a high propensity to be successful in aviation should be encouraged to pursue careers as commercial airline pilots.

4. Scholarships and high-value loans should be developed to aid those individuals who have been evaluated and have demonstrated that they have the potential, but would otherwise be unable to capitalize on their ability and desire to attend a university or college and complete an airline-focused flight training curriculum with the goal of becoming commercial airline pilots.

5. A national aviation industry, university, and government aviation education and raining coalition should be created. This joint coalition would be an on-going forum to define commercial pilot needs, develop training standards, furnish aviation education and training concepts to provide the industry with the best trained and the safest airline pilots in the world.

References

Ballenger, J. (2001, January 8). *Some airlines will lose 25% of work force.* The Public I: An investigative report at the center for public integrity. Retrieved from http: www.publicvi.org/story_or_010501.htm.

Brookfield, S. (1989). Facilitating adult learning. In Merriam and Cunningham (Eds.), *Handbook of adult and continuing education.* San Francisco: Jossey-Bass.

Bruffee, K. (1995, January/February). Sharing our toys: Cooperative learning versus collaborative learning. *Change, 27*, 12-18.

Carretta, T. R. (2000). U.S. Air Force pilot selection and training methods. *Aviation, Space, and Environmental Medicine, 71*(9), 950-956.

Clark, J. B., & Riley, T. L. (2001). Screening EEG in aircrew selection: Clinical aerospace neurology perspective. *Aviation, Space, and Environmental Medicine, 72*(11), 1034-1036.

Cross, K. (1979). Adult learners: Characteristics, needs, and interests. In Peterson, R. and associates (Eds.), *Life learning in America.* San Francisco: Jossey-Bass.

Damos, D. L. (1996). Pilot selection batteries: Shortcomings and perspectives. *TheInternational Journal of Aviation Psychology, 6*(2), 199-209.

Darby, K. (2001a). Pilot Shortages. *Proceedings of 26th Annual FAA Commercial Aviation Forecast Conference*, March 13-14, 2001, Washington, D.C.

Darby, K. (2001b). Aviation Information Resources (Air, Inc.) news release, January 30, 2001. Retrieved from http://www.jet-jobs.com/press%20releases/1.30.01.html.

Darby, K. (2002a). Aviation Information Resources (Air, Inc.) news release, Retrieved April 17, 2002 from http:www.jet-jobs.com/press%20releases/4.17.02.html.

Darby, K. (2002b). Aviation Information Resources (Air, Inc.) news release, Retrieved January 14, 2002 from http:www.jet-jobs.com/press%20releases/1.14.02.html

Darby, K. (2003a). Aviation Information Resources (Air, Inc) news release, Retrieved April 16, 2003 from http://www.jet-jobs.com/press%20releases/4.16.03.html

Darby, K. (2003b). Aviation Information Resources (Air, Inc) news release, Retrieved February 24, 2003 from http://www.jet-jobs.com/press%20releases/2.24.03.html

Darby, K. (2003c). Aviation Information Resources (Air, Inc) news release, Retrieved January 31, 2003 from http://www.jet-jobs.com/press%20releases/1.31.03.html

Department of the Air Force (1999, January). *Rated management primer.* Washington D.C.: Rated Management Task Force.

Federal Aviation Administration (2003). *FAA aerospace forecasts FY 2003-2014.* Retrieved February 2003 from http://apo.faa.gov/foreca02/content_5.htm.

Filipczak, B. (1995, March). Different strokes: Learning in the classroom. *Training, 32*, 43-48.

Karp, M. (1996). *Theoretical aviation training for future airline pilots*. Doctoral Dissertation, Walden University, Minneapolis, MN. UMI 9713644.

Karp, M.R. (2000, October). University aviation education: An integrated model. *Collegiate Aviation Review, 18*, (1).

Karp, M.R. (2003). Regional airline pilot training: An integrated bridge training model. In *Proceedings of the Regional Airline Training Symposium, Orlando, FL, 3-5 March, 2003.*

Karp, M.R., Turney, M.A., Green, M., Sitler, R., Bishop, J., & Niemczyk, M. (2001). Retaining women in collegiate aviation by implementing learning style considerations. *Collegiate Aviation Review, 19,* (1). University Aviation Association Conference, October 24-27, 2001, Nashville, TN.

Karp, M, Condit, D., Nullmeyer, R. (2000, Aug). *F-16 cockpit/crew resource management.* (AL/HR-AZ-TR-1999-0253). Mesa, AZ: Air Force Research Laboratory.

Karp, M.R., McCurry, W.K., McHenry,A.L. & Harms, D (2000). "Aviation education and training for future airline pilots: An integrated process." *Proceedings of the World Air Transportation Training Conference, 26-28 April, 2000, Frankfurt, Germany.*

Klokker, M., Brock-Nannestad, L., Mikines, P., Johnsen, A. F., Garhoj, P., & Vesterhauge, S. (1999). Vestibular autorotation test in primary selection of military student pilots using neural networks. *Aviation, Space, and Environmental Medicine, 70* (10), 959-961.

Leino, T. K., Leppaeluoto, J., Ruokonen, A., & Kuronen, P. (1999). Neuroendocrine responses and psychomotor test results in subjects participating in military pilot selection. *Aviation, Space, and Environmental Medicine, 70* (6), 571-576.

Matthew, R., Cooper, J., Davidson, N., & Hawkes, P. (1995, July/August). Building bridges between cooperative and collaborative learning. *Change, 27,* 35-40.

McCurry, W.K., Karp, M.R., Hayes, P., & Moman, M. (2002b). Aviation education for future regional airline pilots, *Proceedings of Regional Airline Training Symposium (RATS) and Tradeshow, March 3-5, 2002, Orlando, FL.*

McCurry, W.K., Karp, M.R., Hayes, P., & Moman, M. (2002a). Aviation education for future airline pilots. *Proceedings of Conference on Industry and Education Collaboration (CIEC), February 5-8, 2002, Sarasota, FL.*

Retzlaff, P. D., King, R. E., Callister, J. D., Orme, D. R., & Marsh, R. W. (2002). The Armstrong Laboratory aviation personality survey: Development, norming, and validation. *Military-Medicine 167* (12), 1026-1032.

Rodgers, D. A., Covington, E. C., & Jensen, R. S. (1999). Response predispositions and piloting safety. *The International Journal of Aviation Psychology, 9* (1), 73-90.

Taylor, W. W., Moore, S.C., & Roll, C.R. (2000). The Air Force Pilot Shortage. *RAND Corporation.*

Zemke, R. & Zemke, S (1995, June). Adult learning: What do we know for sure? *Training, 32,* 31-40.

An Emerging Profile Of Aviation Management/Leadership Styles and Relational Competencies

Mary N. Kutz
6108 Winfield Drive
Oklahoma City, Oklahoma 73162
E-Mail Address: drkutz@cox.net
Phone: 405-720-9091

and

Timm J. Bliss
312 Willard Hall
Oklahoma State University
Stillwater, Oklahoma 74078
E-Mail address: dallas1@okstate.edu
Phone: 405-744-8064

Abstract

This paper combines a synopsis of the literature with the findings of a qualitative study of the management and leadership styles and relational competencies exhibited by leaders in the field of aviation. Only one study has crossed the many subspecialties in the field of aviation. That study combined with data from available literature offers an emerging profile of management styles and competencies specific to aviation leadership. The self-reported management and leadership styles of aviation leaders highlight the situational aspect of this highly volatile environment but are mindful of the relational or emotional side of leadership even in times of crisis. Although primarily participatory in style, most studies indicate that effective aviation leaders exhibit strong directing behaviors during times of crisis but are conscious of the importance of the emotional and relational aspects of their leadership at all times.

Introduction

The study of aviation management and leadership styles and competencies generally has been confined to military combat crews, commercial transport crews, spaceflight crews, and incident command roles rather than aviation organizational leadership. An analysis of previous aviation leadership research combined with a qualitative study of aviation organizational leaders provides an emerging profile of management styles and competencies specific to aviation leadership. This analysis offers some valuable insight for aviation professionals beginning with a brief overview of leadership in general, followed by a description of some of the competencies that comprise success or failure, and concluding with a summary of the findings.

Contemporary Management and Leadership Research

Goleman, Boyatzis, and McKee (2002, p. 3-5) in *Primal Leadership* identified Emotional Quotient as one of the primary factors in leadership today. Their work and that of many others have recognized the urgency of the *feeling* or *emotional* side of the leadership and management equation in today's environment. Goleman referred to the emotional task of the leader as the "original and the most important act of leadership" in that leaders bring out the best in everyone when they drive emotions positively. When they fail to drive emotions in the right direction, nothing they do will work as well.

Grounded in the findings of neurological research and the study of the brain, the concept of emotional intelligence or Emotional Quotient (EQ) provides a clear picture of the competencies that work and do not work in leadership. Goleman, et al, identified Emotional Intelligence Domains as self-awareness, self-management, social awareness, and relationship management and noted that effective leaders typically exhibit at least one specific competency in each domain (Goleman, Boyatzis, & McKee, 2002, p. 38-29). Recent studies in which they analyzed data from some 500 competency models of global companies revealed that the higher the rank of those considered *star performers* the more emotional intelligence competencies emerged as the reason for their effectiveness. Purely cognitive competencies such as technical expertise surfaced as skills needed to do an average job. Significant strengths in analytic reasoning abilities added just 50 percent more profit and emotional competencies such as self-management added 78 percent, social skills 110 percent greater and those in self-regulation competencies added 39 percent incremental profit.

Transformational leadership involves the process whereby leaders use relational competencies to develop followers into leaders. According to Burns (1978), transformational leadership occurs when people raise each other to higher levels of motivation. Avolio (1999) stated that transformational leadership is *morally uplifting*. Such leaders are deeply trusted and exhibit the moral perspective to

warrant such trust. Their willingness to be vulnerable and to self-sacrifice builds tremendous trust among followers along with identification in their mission.

In a 1997 study by Ross and Offermann, the ability of leaders to revitalize organizations to meet competitive challenges was the basis for a study of transformational leadership, personality attributes, and work group performance. They found that transformational leadership can be predicted on the basis of personality attributes as assessed by subordinates and that an enabling personality profile characterized by encouragement and acceptance was strongly predictive of transformational leadership ratings.

For these reasons, Sergiovanni (1990) and others referred to transformational leaders as moral agents who focus themselves and their followers on achieving higher-level missions resulting in higher levels of trust, loyalty, and performance. Relationships built on trust provide the building blocks for the vital force that differentiates the average team from the highly developed one and the exemplary leader from one who simply gets the job done (Avolio, 1999). In sum, transformational leaders raise the level of identification, moral maturity, and perspective of those they lead. They broaden and enlarge the interests of those they lead and over time, develop their followers into leaders.

Kouzes & Posner (1995) surveyed 60,000 or more respondents from all organizational levels in a variety of public and private organizations and discovered recurring patterns of leadership success, most of which emphasized the emotional or relational aspects of leadership. They found that leaders challenge the process, inspire a shared vision, enable others to act, model the way, and encourage the heart. They further asserted that love – being in love with leading, with the people who do the work, with what their organizations produce, and with those who honor the organization by using its work – may be the best kept leadership secret of all. Thus, leadership is an affair of the heart, not of the head.

Retired General Norman Schwarzkopf believed in love. During an interview by Barbara Walters, she asked him how he would like to be remembered. He replied, "That he loved his family. That he loved his troops. And that they loved him (ABC News, 1991)." The famous coach of the Green Bay Packers, Vince Lombardi, emphasized love as well. He made the following remarks in a speech before the American Management Association: "Mental toughness is humility, simplicity, Spartanism. And one other, love. I don't necessarily have to like my associates, but as a person, I must love them. Love is loyalty. Love is team work. Love respects the dignity of the individual. Heartpower is the strength of your corporation" (Peters & Austin, 1985, p. 341). When leaders encourage others, through recognition and celebration, they inspire them with courage - with heart.

All of these studies stressed the emotional and relational competencies of leaders as major contributors to effectiveness in the role. The following findings

in the aviation field are consistent with these previously discussed leadership styles and competencies.

Management and Leadership Styles in Aviation

In a 1995 issue of *Aviation, Space and Environmental Issues*, Nichols and Penwell described their literature review of leader characteristics and outcomes from four environments: aviation, submersibles, polar stations and expeditions. Their aviation leadership studies involved combat bomber crews and more recently commercial transport crews.

Their research explored 23 sources and their findings indicated that aviation leaders shared a common core of personal traits and leadership styles. In comparing their findings across sources, only two contradictions surfaced with a negative association between leader consideration and leadership effectiveness. Leadership consideration (friendship, trust, warmth) or the relational side of leadership was positively associated with leader effectiveness in every study except with World War II bomber commanders. In wartime, initiating structure or finding ways to get the job done is what counts the most and effectiveness is based on a crew's ability to hit the target, whatever the sacrifice. However, even during bombing runs and critical incidents where an air crew was down, the survival of the crew often hinged on the ability of the leader to use relational skills and keep the crew focused on the task before them.

The profile that emerged from their research was that of a leader who worked hard to achieve mission objectives, was optimistic, commanded the respect of the crew, used participative decision-making but took charge in critical situations, was sensitive to crew members feelings and made them feel valued for their expertise and personal qualities and maintained group harmony and cohesion. (Nichols & Penwell, 1995, p. 63)

Flin (1997) described attributes of a successful incident commander in terms of personality characteristics as: leadership ability that inspires trust and commands respect; a stable personality that demonstrates emotional stability, maturity, and steadiness; and the ability to formulate and implement decisions under pressure and determine when to use authoritative or consultative decision making style. These attributes also parallel emergency aviation services personnel, e.g. Air Force and international commercial airlines.

Aviation Organizational Leadership

Although very little research has been done on aviation organizational leadership styles, a cursory review of the literature and an analysis of commercial airline successes and failures provide some interesting insights.

Students of an aviation leadership graduate class at Oklahoma State University recently were assigned to research the literature pertinent to the history of

the major air carriers for insights on the leadership of those airlines that failed as well as those which still exist. They were asked to analyze the style of all of the leaders of each company and determine what, if any, contribution each of the leaders had made to the success or failure of their companies. An additional part of their learning assignment was an analysis of each airline to determine whether the "Southwest style" might have made a difference in the success or failure of the company, even in the environment of aviation during the era the companies existed. Southwest Airlines, which has been profitable for over thirty years and considered by most to be the world's most successful airline, was used as the standard-bearer of effective leadership for obvious reasons (Gittell, 2003). Their emphasis on the emotional aspect of leadership is readily apparent throughout their organizational history.

The outcome of this class assignment was that, in literally every case, there appeared to be a direct causal relationship between the leadership style and the ultimate success or failure of the company. Almost without exception, those companies that went out of business began their trend downward with the advent of leadership that struggled with the relational side of the leadership equation. Frequently those companies could not recover even when replaced with strong leadership. The consensus of the students was that leadership not only was related to success, but also was causal in that most, if not all, cases of failure that occurred when a manager or leader who was weak in relational competencies started the company down a path from which they could never fully recover. The consensus of the students was that the Southwest style is a product of a culture that has been carefully nourished over time to meet the needs of the culture of today's workforce. It would have represented a dramatic shift for the workforce of an earlier era and might have required gradual progression; but even a benevolent dictator, who offers some concern for the emotional or relational side of leadership, might have had a better chance for leading a company to success than some of the leaders did.

Self-Reported Management and Leadership Styles of Aviation Leaders

Bennis and Thomas (2002, p. 123) reminded us that regardless of the era, leaders share certain commonalities. "Geeks" (leaders age 35 and under) belong to a digital era of flat organizations and "Geezers" (leaders or grandparents of Geeks) are from an analog era of organizational hierarchy and chain of command leadership. In spite of the differences in eras, Bennis and Thomas were able to identify a Leadership Development Model of leadership competencies that crosses all generations. Those competencies included *adaptive capacity* (hardiness, creativity); *engaging others by creating shared meaning* (empathy, obsessive communication); *voice* (purpose, self-awareness, EQ); and *integrity* (competence, ambition, moral compass). They also pointed out that we learn as much about leadership from failures as from successes. Many of the leadership failures in the airline industry demonstrated a glaring deficit in those competencies related to the emotional side of leadership.

As demonstrated by the previously discussed work of Bennis and Thomas, one of the best ways to learn about leaders is to ask the leaders. One of the most recent studies that crossed occupational specialties of aviation leaders utilized interviews to surface valuable data about management and leadership styles. In a small 1998 qualitative study of the Characteristics of Successful Aviation Leaders in Oklahoma, eighteen aviation leaders from a variety of different aviation specialties were interviewed regarding their management and leadership styles. Participants chosen for the study were defined as successful aviation leaders if they had achieved the top ranked position in an aviation organization or had achieved a position of influence in the community as an aviation leader. Their organizations ranged in size from three to 20,000 employees and the scope of their responsibilities ranged from local to international (Kutz, 1998).

Each of the leaders in the study was asked a series of questions pertinent to their perspectives of their own aviation management and leadership styles. One of the questions requested that the leaders describe their management or leadership style. A follow-up question was sometimes asked to spur thought. That question offered descriptors and asked if the leaders saw themselves as nurturers, protectors, problem solvers, pushers or some other appropriate term.

All of the aviation leaders interviewed mentioned multiple management and leadership styles and emphasized the relational side of the equation regardless of the style. Most described themselves as *participatory* managers with a back-up style of *directing*. However, they used relational terms to describe their style even when it became necessary to resort to *directing*. The term most frequently used by participants to describe their own style was that of *facilitator;* and most of the terms addressed soft skills or interpersonal skills such as *motivator, energizer, coach, mover, communicator, rewarder, protector, and teacher or trainer.* Such descriptors as *honesty, openness, sensitivity, sincerity, warmth, listener, caring, trusting* and *feeling* were often used to describe the relational aspects of their primary leadership characteristics associated with their style.

One of the leaders offered a unique example of the emotional quotient at work under unique circumstances. He described himself as a *situational* manager who leads by providing a bit of direction and vision but acknowledged that sometimes direction IS *direction* in an emergency and that direction can be an important part of the emotional quotient of leadership. Emotions run high during crisis and the ability of the leader to diagnose and deal with those powerful feelings and emotions associated with life and death become paramount not only for the leader but for the followers. Sometimes that means directing. Even in an aircraft accident with bodies all around and people upset, someone has to take charge. He offered an example of a circumstance when direction became necessary to deal with the emotional quotient after the bombing of the Murrah Building in Oklahoma City. His managers later described his firm, motivational talk soon after the incident as one designed to remind them that they were selected to lead the people regardless of their own personal emotions or grief.

According to this leader, when things settle into a state of normalcy, the emotional side of leadership is no less important in that it is once again important to listen and provide opportunity for participation and even failure. Regardless of the demands of the situation and the management style in use in a given situation, it was readily apparent that the emotional side of leadership was always a consideration.

Another aviation leader described his personal style as *democratic* leadership or *dictatorial* depending on the circumstances. He elaborated by describing himself as very demanding as far as standards are concerned. He believed in the importance of doing something right or not at all, which meant striving for excellence and not settling for second best. He believed that the relational aspect of management and ability to obtain buy-in from those who reported directly to him was an important part of achieving that excellence. If that was not possible, he began "dealing in the dictatorial sector of the spectrum" (Kutz, 1998) ever mindful of the impact of his actions on relationships in the organization.

One of the leaders in the Kutz study who described himself as *participative* recognized that he was surrounded by capable adults who do not need to be treated as children. He described his role as a coach who provides the training and resources then gets out of the way so that people can do their jobs. He stressed the importance of family life and emphasized that people who put in eight-hour days must be working smarter than those who put in 12 hours days.

Still another of the leaders described his style based on feedback received from three Blake and Mouton Managerial Grid Seminars. He described his feedback as 9,9 or *participative* and *team-oriented* with a backup style of 9,1 or *directive* and *production oriented*. He expressed concern with a tendency to run out of patience too soon and become directing. He recognized that after a crisis when things return to normal, an autocratic style can damage an organization and that the relational side of leadership cannot be neglected regardless of the circumstances.

An interesting outcome of all of the interviews was that none of the leaders described a single dominant style for all occasions. Although they most frequently described their preferred style as participatory, which implies concern for relationships and emotions, they often resorted to a directing style in times of crisis. Yet, even when their preferred management style was *directing*, they referred to a participative back-up style. All of the leaders repeatedly expressed their concerns for the relational side of leadership. Some of the leaders were concerned enough for their own relationships with their employees to suggest that a parallel study should be conducted asking their employees the same questions to see if the perspectives of their leaders were consistent with the leaders' perspectives of their own styles.

Leader Self-Analysis of Aviation Leadership Compared
To Other Leadership Roles

When asked to define successful aviation leadership and distinguish the difference between aviation leadership and other leadership roles, the answers most frequently addressed the importance of high standards in getting the job done and achieving the mission; tapping the passion and love for aviation to achieve maximum performance; and emphasizing the ability to balance innovativeness with risk taking. Although the interviewees were divided equally in their opinions of the uniqueness of aviation leadership, they were not divided on the importance that emotions and relationships play in a high-risk, volatile environment such as aviation. Approximately 50 percent saw no difference in aviation leadership and that of other organizations, but cited the importance of mission, influence, motivation, and relational skills in achieving success. Some of the ones who saw no difference qualified their statements by mentioning the importance of the safety factor, the risk, and the flamboyance (the perceived love) for the business and the need to be cognizant of the emotional aspect of each of these factors.

Those leaders in the Kutz study, who did perceive the aviation leadership role as unique, cited such reasons as the volatility, capital intensive, and competitive nature of the aviation environment; the glass ceiling based on ratio of women to men in the industry; a perceived greater emphasis on quality because of the unforgiving nature of the profession; the more structured and regulated aviation environment; the critical importance of being futuristic and visionary in such a volatile industry; the impact of the overall mission on the public; the technical requirements, and the overall love for aviation – which speaks to the importance of the emotional or relational side of leadership.

Overall, the general profile of aviation leadership that emerged from all of these discussions with the Oklahoma leaders was one of flexibility of style with the dominant style of participation and consideration supported by ability to transition quickly to *directing* but with a continual awareness of the relational needs of the organization. By providing choice and flexibility, aviation leaders are major contributors to enriching a sense of personal well-being and increasing performance (Karasek & Theorell, 1990). If leaders want to reach higher levels of performance and less dependence, they must be proactive in designing tasks that allow people discretion and choice. Having flexibility means being liberated from a standard – *non-transforming* – set of rules, procedures, or schedules as circumstances permit with a continual awareness not only of the mission but also of the needs of the people responsible for achieving the mission.

Summary of Findings

Leaders are people who understand that without vision, there is no leadership; without honesty and trust and respect for the relational aspect of leadership, there is no communication; without communication and a common core of

values, there is no performance; and without the confidence and commitment to reach out and try new things, there is no risk taking or long-term success.

In this review of the literature, although leaders in different occupational specialties may paint a self-portrait using a variety of different "leadership style" descriptors, a number of consistencies actually surfaced among these different occupational specialties. The transformational leaders capable of transforming organizations, regardless of the different environments, share some commonalities of style and competence that can be labeled consideration, influence, emotional intelligence or any of a variety of descriptors that generally result in a more participatory management style.

Although the perceptions of aviation leaders may vary somewhat it terms of the unique qualities necessary to survive in an aviation leadership environment, all expressed attributes and commonalities of style with leaders in different occupational specialties. There also were striking similarities across the differing occupations pertaining to backup styles that are used when circumstances necessitate speed of *direction* and decision-making. The implications of the research findings indicated the importance of flexibility in leadership and the development of a range of styles that permit rapid decisive response as needed.

Given the complexities of leadership in the 21st century, the authors believe the impact of leadership styles and power issues will demand continuous research in understanding the emotional side of management and leadership in all occupational specialties, particularly aviation. One thing is certain, the emotional or relational side of management and leadership cannot be neglected in the global world of aviation where people use aircraft for every purpose from commerce to weapons and where responsibilities range from recreational to monetary to life and death issues.

References

ABC News. (1991, March). *20/20: Barbara Walters' interview with General Norman Schwarzkopf*. Burbank, CA: ABC, Inc.

Avolio, B. J. (1999). *Full leadership development: Building the vital forces in organizations*. Thousand Oaks, CA: Sage Publications.

Bennis, W. G. & Thomas, R. J. (2002). *Geeks & geezers, how era, values, and defining moments shape leaders*. Boston: Harvard Business School Press.

Burns, J. M. (1978). *Leadership*. New York, NY. Free Press.

Flin, R. (1997). *Sitting in the hot seat*. N Y: John Wiley & Sons.

Gittell, J. H. (2003). *The Southwest Airlines way, using the power of relationships to achieve high performance*. NY: McGraw-Hill.

Goleman, D., Boyatzis, R. & McKee, A.. (2002). *Primal leadership, realizing the power of emotional intelligence*. Boston, MA: Harvard Business School Press.

Karasek, R. & Theorell, T. (1990). *Healthy work: Stress, productivity, and the reconstruction of working life.* New York, NY: Basic Books.

Kouzes, J. M. & Posner, B. Z. (1995). *The leadership challenge.* San Francisco, CA: Jossey-Bass Publishers.

Kutz, M. N. (1998). *Characteristics of successful aviation leaders of Oklahoma.* Unpublished doctoral dissertation, Oklahoma State University, Stillwater, OK.

Nicholas, J. M. & Penwell, L. W. (1995). A proposed profile of the effective leader in human spaceflight based on findings from analog environments. *Aviation Space & Environmental Medicine.* Jan.-Jun 1995. 66(1), p. 63-72.

Peters, T. & Austin, N. (1985). *A passion for excellence, the leadership difference.* New York: Warner Books Edition.

Ross, S. M. & Offermann, L. R. (1997). Transformational leaders: Measurement of personality attributes and work group performance. *Personality and SocialPsychology Bulletin.* October 1997. 23(10), 1078-1086.

Sergiovanni, T. J. (1990). *Value-added leadership: How to get extraordinary results in schools.* Orlando, FL: Harcourt Brace.

American Airlines' Attempt to Implement Rolling Hubs

Karrie Shank and Mark Sherman
Ohio University
Department of Aviation
Ohio University Airport
751 Columbia Road
Albany, OH 45710

Abstract
Since the Airline Deregulation Act of 1978, major airlines have abandoned point-to-point service adopting the hub and spoke system. "Hub depeaking, which spreads flights out during the day instead of arranging them in the peaked connecting banks previously used, aims to promote higher aircraft, employee, and gate utilization without making large schedule cuts" (Goedeking, 2003, p. 93). With American Airlines in turmoil after September 11, 2001, combined with the loss of business travelers, American Airlines began experimenting with a "rolling hub" concept, also known as depeaked hubs. Chicago O'Hare was the first to utilize this experiment in June 2002, followed six months later at Dallas/Fort Worth. American has not attempted to implement rolling hub operations at their St. Louis, Miami, or San Juan hubs due to a decline in passenger revenue that necessitated parking aircraft and furloughing crews. Depeaking was estimated to create the equivalent of 17 new aircraft saving the company $1.3 billion of future capital spending.

Introduction

The hub-and-spoke system came into existence after the federal government deregulated the airlines in 1978, removing federal control over airlines' fares and services. Airlines could now set their own prices and make their own decisions regarding many aspects of operation including routes, in-flight service, corporate attitude, and quality of service they were willing to provide.

Before the Airline Deregulation Act of 1978, the Civil Aeronautics Board (CAB) controlled all airline operations – everything from ticket price to route structure. Some CAB mandated routes were not profitable, but the CAB awarded lucrative routes to offset losses. In 1977, economist Alfred Kahn assumed CAB Chairman duties and quickly realized that if a deregulated airline industry existed, competition among carriers would expedite driving down the fares. With deregulation, more companies would be able to enter the market serving smaller communities, and it would be better for all parties involved.

After deregulation, airlines established 32 hub complexes as part of the hub-and-spoke system to save capital; however, they were costly to establish (Ott, 2002). "The hub-and-spoke system has been widely blamed for the gridlock that often develops when flights are delayed by weather or other reasons in one part of the country. But it is also very inefficient," says Gellman, "because it requires large numbers of ground workers to service arriving and departing flights simultaneously" (Chandler, 2002, p. 1). An airline hub needs additional terminals and ground crews, which means increased payroll and decreased profit levels. Another problem encountered with hubs is heavy aircraft and passenger congestion occurring at peak times when many business travelers fly. Major schedule disruptions frequently occur by a single aircraft experiencing mechanical problems or weather systems in one part of the country that in turn affect flights scheduled to depart airports many miles away.

Effects of Deregulation

Deregulation has had both positive and negative influences upon the airline industry, summarized as follows:

Table 1

Deregulation's Positive and Negative Influences (Kane, 2003)

POSITIVE	NEGATIVE
Airline fares have increased at a rate lower than that of inflation	Fares initially decreased, but have risen dramatically as industry concentration has increased
More discount fares available	Concentration of hubs with little competition
Fares to small cities are only 3% higher than at major airports	Nonstop flights between many destinations have disappeared as a result of hub and spoke systems
Airlines & passengers have saved large sums due to deregulation	Rural areas have been adversely affected by flight schedule reduction and higher fares
Airlines have been innovative in transition to the deregulation environment	Fare system is confusing. Raises traveler equity questions
Hub & Spoke structure has led to significant increase in flight frequencies	Airline industry has not paid adequate attention to consumer issues
Average airline industry earnings showed improvement prior to onset of 1990's recession	Greater industry concentration as major airlines fail and are not replaced

Deregulation had several positive impacts upon the industry. Airline fares increased at a rate below that of inflation. This resulted in lower fares, which in turn attracted more passengers and increased revenue. More discount fares became available and this, along with fares to smaller cities being only 3% higher than major airports, made flying more economical for passengers who originally may have not elected to fly. Deregulation gave the airlines more flexibility in their planning, enabling them to save money because there was no CAB forcing them to keep unprofitable routes and requiring approval of every minute decision. Being able to act upon decisions without the CAB's approval allowed airlines to be more innovative in their attempts to increase passenger shares and revenue. The airlines were able react to passenger demands and route changes in a timely manner. The hub and spoke system created new opportunities for flight connections that the linear system could not. For every flight added in a linear system, one connection was created, but under hub and spoke organization, the connections increased exponentially for every flight added. All of these factors resulted in increased revenue in the years before the 1990's recession. This increase was evidenced by orders being placed for new aircraft and increases in passenger counts.

While there were positive impacts of deregulation, it was not an entirely beneficial transition. While fares initially were increasing at a rate less than that of inflation, this trend did not continue indefinitely. As airlines were able to act upon their own decisions, certain airlines gained a larger presence at certain airports, which became their hubs. As airlines became dominant at their hub airports, a sort of monopoly was created. This monopoly, coupled with the lack of competition, both at that airport and system wide, drove fares artificially high. In addition to driving fares higher than necessary, the fare system has become increasingly complicated. Fares on one flight vary, raising the question of whether one passenger is worth more than the other, and will receive better service, while one is inconvenienced. These feelings may have driven some passengers to find alternate methods of transportation. The increases in fares, along with cancellations of flights to smaller cities as airlines did not feel compelled to respond to passenger demands due to lack of competition has driven away passengers who felt inconvenienced by the changes. Overall, lack of competition has caused the airlines to have a certain disregard for passenger issues, and this trend increases as smaller airlines fail and are not replaced, resulting in even greater industry concentration.

The effects hubs impose on the flying public and air traffic control are enormous; "281 locations or less than 2% of all airports in the United States, handle virtually all of the airline passengers. The top 20 airports account for almost two-thirds of all enplanements, and the top 10 account for 40 %. Close to one-quarter of all airplane passengers board their flights at one of just five airports: Chicago O'Hare, Atlanta Hartsfield, Los Angeles, Dallas-Fort Worth, and New York Kennedy" (Wells, 1986, p. 35-36).

Under the rolling hub concept, airlines reduce aircraft idle time by decreasing the number of flights scheduled around crucial peak rush hours. "Arrival and departure windows of one hour allow the flexibility to slide flights earlier or later to depeak the spokes while maintaining connections. De-peaking eliminates "dead-time" at the hub by operating no more than one arrival and one departure per minute throughout the day" (Steering Committee, 2003, p. 3).

In 2001, American Airlines took a new look at their hub airports, looking for a more efficient way to handle daily traffic. Due to recent fallout of passengers, the airline was loosing considerable amounts of money. American had to stop the red ink that occurred due to the acquisition of T WA and the post 9/11 environment. The result was the concept of "depeaking" or "rolling hubs," or simply spreading flights out more evenly over the course of the day. By using rolling hubs, American hoped to achieve increased aircraft utilization, decreased ground time, more departures, fewer gates, increased revenue, and decreased costs.

Before September 11, 2001, American clustered flights around peak flying hours in hub airports. George Hamlin, Senior Vice President at Global Aviation Associates Ltd., stated rolling hubs should enhance productivity and produce significant cost savings. "Aircraft may push back 1-2 minutes apart, but they show up on runways and they bunch up. In a rolling hub, flights come and go without having to be in a bank. Before this, American was paying people [and for aircraft] to sit still, and now they are paying them to move" (Ott, 2003a, p. 53). "The message is clear that to stay in business for the long term, profitability is more important than market share" (Ott, 2003b, p. 22).

After September 11, 2001, the entire industry changed. Most airlines were dealing with high debt and decreased load factors, (The percentage of seats or freight capacity utilized. Seat load factor is derived by revenue passenger miles divided by available seat miles) which are now starting to return to normal levels after two years.

Table 2

Percentage Changes In Traffic, Capacity, And Load Factor (Bond, 2003a)

AIRLINE	Traffic July 03 Vs July 02	Traffic July 03 Vs July 01	Capacity July 03 Vs July 02	Capacity July 03 Vs July 01	Load Factor July 03 Vs July 02	Load Factor July 03 Vs July 01
American	-.40%	-10.10%	-6.90%	-14.80%	+5.3 pts	+4.3 pts
United	-5.70%	-17.30%	-12.2%	-23.70%	+5.6 pts	+6.4 pts
Delta	-2.70%	-9.10%	-8.5%	-14.50%	+4.9 pts	+5.0 pts
Northwest	-3.90%	-11.60%	-7.50%	-15.30%	+3.2 pts	+3.6 pts
Continental	8.90%	0.50%	2.60%	-6.60%	+4.9 pts	+5.9 pts
U S .Airways	-3.70%	-21.20%	-9.70%	-25.80%	+5.0 pts	+4.7 pts
Southwest	9.80%	10.3%	3.60%	7.80%	+4.3 pts	+1.7 pts

Before September 11, 2001, full-fare travelers started to disappear.American announced a plan to streamline their fleet, thereby recouping capital with less maintenance by retiring fleets or subfleets, deferring new aircraft acquisitions, and eliminating first class seating to selected markets (Sofradzija, 2002). After September 11, 2001, fare prices changed. The average cost per enplanement increased anywhere from 4.5% at Cleveland Hopkins to 94.5% in San Francisco (Ott, 2002). Only a few rare cities experienced decreases in average enplanement costs. With 477 flights per day at one hub airport alone, costs add up rapidly.American expected to loose $7 billion for the second year in a row (Flint, 2002).

American Airlines

From a management standpoint, depeaking had the potential to create several money-saving opportunities. Rolling hubs could allow increased aircraft utilization, allowing for added flights, or decreased fleet composition without incurring flight cancellations. By not banking flights at peak times, American aircraft could spend less time sitting on the taxiway, waiting for takeoff, resulting in better on-time performance and implementation of a standard turnaround time. Gate rentals were saved by returning unused gates to the airport authority, thereby requiring fewer customer service employees and less ground equipment. There was a risk of loosing customers to other airlines but the potential gains outweighed risks.

Chicago O'Hare

In June 2002, American began utilizing rolling hubs at their omnidirectional Chicago O'Hare hub, serving both east/west and north/south traffic. American initially chose O'Hare since Dallas/Fort Worth is America's largest hub with 675 flights every weekday. Management considered this change the biggest gamble taken in over two decades, expecting to save $100 million a year in fuel, facility charges, payroll, and provide relief at spoke airports (Ott, 2003a).

Ordinarily, such a move would not have been considered such a risk, had it happened in another industry; however, several characteristics of the airlines make it a larger gamble than it seems at first glance. In the airline industry,a decision such as this cannot be made on a small scale. Depeaking was an all-or-nothing choice. If it was going to be done, the change had to be made simultaneously to all flights arriving and departing at that airport. Unlike smaller businesses, where unwise decisions may cause small losses, had passengers not approved of the changes and decided to travel on other airlines, the losses had the potential to be of a much larger scale. Decisions within the airline industry also have a history of being made with close regard to what other airlines are doing or will do. Since depeaking had never been done, like many other decisions, it was considered to be a bad idea since not everyone was doing it.

Initial results appeared positive. Except for hourly flights between O'Hare and New York's LaGuardia, American adjusted practically all 333 mainline flights and 180 Eagle flights with a 5% labor reduction (Ott, 2003a). Average connect times (The legal minimum time necessary to change planes at a given airport usually considered to be 45 minutes) increased by an average of 10 minutes; ticket counter lines were consistent, but shorter. Turnaround times slightly increased, aircraft spent less time sitting on the taxiway waiting to depart, overall ground time at both hub and spoke airports decreased, and on-time performance improved. Departing passengers increased slightly while connecting traffic saw a minor drop. American also implemented a 42-minute turnaround requirement for each Chicago-served spoke city (Ott, 2002).

Table 3

Analysis of Flights Before and After Depeaking O'Hare (Marta, 2002)

DEPARTURES PER HOUR	BEFORE DEPEAKING	AFTER DEPEAKING
11:00 A.M. – 12:00 P.M.	34	20
12:00 P.M. – 1:00 P.M.	4	22

Flights before depeaking show 34 departures from 11:00 A.M. through 12:00 P.M., and only four from 12:00 P.M. to 1:00 P.M. Under rolling hubs, there are 20 departures from 11:00 A.M. through 12:00 P.M. and 22 departures from 12:00 P.M. to 1:00 P.M. American was able to operate their full schedule with four fewer gates, five less airplanes, less ground crew, and fewer lost bags (Marta, 2002). Taxi times decreased by 1.5 minutes, better than predicted (Torbensen, 2002). Average block times (The amount of elapsed time between an aircraft leaving the departure ramp for the purpose of flight and its reaching the arrival ramp at the end of the flight. From the time the parking brake is released for pushback to the time the parking brake is set upon arrival) were down by ten minutes; worth at least $4.5 million, but still higher than levels United had on common routes. Donald Carty, former Chairman and CEO of American Airlines, said, "Our Chicago experience has improved customer service, reduced costs, improved productivity and allowed us to fly the same schedule with the equivalent of five fewer aircraft and four fewer gates" (Ackerman, 2002, p. 2). "We expect the DFW and spoke de-peak to allow us to fly an equivalent schedule with 11 fewer aircraft, with an as yet undetermined number of gates saved as well" (Magers, 2002, p. 1). Don Casey, American Airlines' Managing Director for Scheduling, estimated savings of $100 Million a year from reduced cost for fuel, facilities, and personnel. "One minute of block time at American Airline's Chicago hub was worth $4.5 - 5 million, and 50% more at American's Dallas hub due to its larger size. With reduced block time, American gained efficiency as aircraft burned less fuel. Personnel cost, especially for pilots (which is based on scheduled or actual block time, whichever is larger) has been a major part of the reduction," (Ott, 2003a, p. 53). However, market share in Chicago shows Ameri-

can lost four market share points while United moved quickly to take advantage of runway and airspace availability. Overall, American lost part of its market share at Chicago, but the main goal of depeaking was not to increase market share, but to increase the percentage of seats filled on each flight arriving and departing the airport.

Dallas/Fort Worth

After the general success at O'Hare, American started to depeak their Dallas/ Fort Worth hub in November 2002. The result was a reduction of four gates, all mainline flights consolidated into Terminals A and C ("American Airlines, American Eagle," 2002) with a cessation of mainline Terminal B operations at an estimated savings of $4.5 Million compared to expenditures prior to depeaking ("American to Save," 2002). The changes enabled American to continue their established schedule with nine fewer mainline aircraft, two less regional aircraft, as well as reductions in ground crew, gates, and ground equipment (Ott, 2003a).

Table 4
Changes to American's Operations at DFW (Goedeking, 2003)

	BEFORE DEPEAKING	DEPEAKED
Max outbound flights per 15 min	48	20
Max inbound flight per 15 min	36	21
Max movements per 15 min	50	37
Average block hour (outbound)	2:22	2:12

Maximum number of scheduled movements per 15 minutes decreased from 50 to 37 resulting in a loss of connectivity. Connectivity quotient decreased from 2.14 to 1.71 but allowed American to make 71% more connections than would be delivered by a random distribution. Average block time also decreased, by 10 minutes. Banks are no longer omnidirectional, but have a pronounced directional structure. Delta, American's rival at Dallas, is not shifting to a rolling hub concept, but is adding two additional banks. American lost 1% market share primarily due to little competition (Goedeking, 2003).

St. Louis

"Using traditional Transportation Department measures, St. Louis is the most efficient American Airlines hub, and it's among the highest rated in the industry" (Wilson, 2003a, p. E9). After depeaking O'Hare and Dallas/Fort Worth, American had to decide if depeaking Saint Louis and Miami would be beneficial. The difficulties would be at Saint Louis-Lambert, a predominantly east/west hub which does not receive significant Chicago traffic. American considered two options for the Lambert hub acquired with the assets of Trans World Airline (TWA). The first would have dropped all hub operations in St. Louis and turned

this facility into a spoke for both Dallas/Fort Worth and Chicago. The second option was to reduce all operations at Dallas, St. Louis, and Chicago. American ultimately chose a middle ground in dealing with Lambert Airport (Bond, 2003b).

After acquiring T WA's assets in 2001, American has essentially killed the T WA St. Louis Hub. In November 2003, American reduced Lambert operations from 417 flights serving 94 markets per day to 207 flights serving 68 markets (Kittle, 2003). Of its remaining 207 daily departures, 53 are mainline jets operated by American; regional airline partners, American Connection and American Eagle, fly the remaining 154 flights. American also cancelled nonstop service to 27 cities from Lambert, 31 gates closed, leaving American with only 18 gates at Lambert. The new schedule was a drastic reduction from the daily 522 nonstop departures the airline and its affiliates offered three months after buying Trans World Airlines' assets out of bankruptcy in 2001 (Leiser, 2003). Dan Garton, Executive Vice President of Marketing for American stressed that American is not abandoning St. Louis; it intends to maintain Lambert as a domicile for both pilots and flight attendants (Wilson, 2003b)

American's Other Hubs

American has additional hubs located in Miami and San Juan. Miami appeared able to support depeaking, but would be difficult considering flight length and direction. In the years after November 11, 2001, Miami and San Juan experienced a decline in passenger revenue, necessitating parking aircraft, and in turn furloughing crews. The decreased number of aircraft and flight numbers operating out of these airports resulted in increased load factors on the remaining flights, essentially accomplishing the same thing that the rescheduling at Chicago O'Hare and Dallas/Fort Worth had. The loss of passenger revenue essentially depeaked both airports without assistance from the company. Due to the downturn in the aviation industry, only Chicago and Dallas/Fort Worth have been depeaked by the company.

Conclusion

There were a few losses because of depeaking. United Airlines, American's main competitor, recorded an increase in passengers, but it did not seem to outweigh gains for American. American Airlines is not abandoning hubs, just reorganizing the system. One major difference between traditional hub-and-spoke system and the rolling hub system is flow of airplanes. In a hub-and-spoke system, planes arrive in waves of up to 50 within minutes of each other, and depart the same way. With a rolling hub system, planes come in and out in a steady flow throughout the day. "While aircraft ground time at the hub doesn't change much, planes can depart 'spoke' airports as quickly as they can be reloaded. In a traditional hub-and-spoke operation, planes sometimes remain at spokes for long periods, timing their return to the hub to be a part of a bank of arriving flights" (Reed, 2002, p. B2). Casey estimated gate reductions from 286 to 252 across the network (Ott, 2003a).

Competitors will monitor American Airlines' operations under the rolling hub strategy for analysis. Overall, American Airlines felt that depeaking accommodated all passengers better than the old system. American also is changing the level of service it offers to a number of cities. Some travelers will fly on smaller, regional jets rather than big airliners.

Other airlines will watch American closely during this experiment and consider remodeling their own schedules if American is successful. Nobody had attempted to depeak such a wide range of airports before. Some airlines such as Continental implemented the rolling hub concept at their Newark hub and talked about implementation at their Houston hub. For the first time, an airline such as American tried implementation at all their hubs. True, only Chicago and Dallas were actually implemented; St. Louis and Miami were depeaked by natural occurrence with aircraft being parked and crews furloughed. While different approaches may have been taken at these different airports, the result was the same. Some continue to speculate that eventually depeaking will prove to be American Airlines' downfall, while others predict that in the future, airlines will be forced into bankruptcy though failure to change their scheduling policies. "The depeaking and fleet efficiencies will create the equivalent of 17 new aircraft, saving the company $1.3 billion of capital spending in the future" (Ott, 2003, p. 23).

References

Ackerman, H. R. (2002, September 1). American to revamp DFW operations. *Airport Report, 47* (17), 1-2.

American Airlines, American Eagle shift terminals at DFW to enhance customer convenience. (2002, October 22) *PR Newswire Association, Inc* .

American to save $4.5 million by switching terminals at DFW. (2002, October 23). *Aviation Daily*

Bond, D. (2003, August 11). Traffic on a treadmill. *Aviation Week & Space Technology.*

Bond, D. (2003, July 21). AMR's St. Louis blues. *Aviation Week & Space Technology* .

Chandler, S. (2002, Dec. 8). "Trying to change course". *Chicago Tribune* .

Flint, P. (2002, November). No peaking. *Air Transport World* .

Goedeking, P. (2003, September 1). "Breaking the bank; A close look at the mechanics of the rolling hub system reveals how to minimize connectivity loss while improving operations". *Airline Business, 93* .

Kane, R. M. (2003). *Air transportation*. Dubuque, Iowa: Kendall Hunt Publishing Company.

Leiser, K. and Wilson, C. (2003, October 26). American cuts at Lambert hit home this week. *St. Louis Post-Dispatch* .

Kittle, M. D. (2003, July 19). Airport hopes to add St. Louis; Other facilities are
   losing flights as American plans to cut service. *Telegraph Herald Dubuque*,
   IA.

Magers, P. (2002, August 14). American to cut 7,000 jobs, ground planes. *United
   Press International*.

Marta, S. (2002, August 15). American Airlines is banking on "rolling hub" sys-
   tem. *Knight Ridder Tribune News*.

Ott, J. (2002, September 2). American takes a risk with rolling hub. *Aviation
   Week & Space Technology.*

Ott, J. (2003, February 16). 'De-peaking' American hubs provides network ben-
   efits. *Aviation Week and Space Technology.*

Ott, J. (2003, August 19). Top U.S. majors shift into survival mode. *Aviation
   Week & Space Technology*.

Reed, D. (2002, August 8). "American Airlines to try rolling hubs". *USA Today*.

Sofradzija, O. (2002, August 14). Airline carrier's cost-cutting probably won't
   affect Peoria—American to trim 7,000 jobs, remove seats, jet fleet. *Journal
   Star.*

Steering Committee. (2003, May 8). Schedule de-peaking. *American Airlines
   Now*.

Torbenson, E. (2002, November 1). American Airlines to roll out new flight plan.
   Knight Ridder Tribune News.

Wells, A. T. (2000). *Airport Planning & Management* (4th ed.). Blacklick, OH:
   McGraw-Hill.

Wilson, C. (2003, February 9). A cloudy picture for St. Louis; status of the
   Lambert Field hub. *St. Louis Post-Dispatch*.

Wilson, C. (2003, July 17). AA slashes operations here; cuts are all about money,
   airline chief tells analysts. *St. Louis Post-Dispatch*.

Book Reviews

Air Rage – The Underestimated Safety Risk
by Angela Dahlberg

Reviewed by
Deak Arch

and

Mark Sherman
Ohio University

"Air rage" is a term that has been coined by the media within the last few years to describe a situation where passenger misconduct occurs aboard an aircraft. As of yet, air rage is not a Federal Aviation Administration (FAA) defined term. However, passenger misconduct is clearly defined, according to Title 14 of the Code of Federal Regulations Part 91.11, as any threatening, intimidating, or interfering act with a crew member. In addition, the FAA further addressed misconduct of passengers in their Advisory Circular 120-65 by defining any interference with crew members' work performance as misconduct. It is the continued act of misconduct by a passenger, despite remediation by crew members, which may result in federal prosecution.

According to Angela Dahlberg in her book, *Air Rage – The Underestimated Safety Risk*, air rage is defined loosely as anything ranging from verbal abuse to assault in the aircraft cabin. The author utilized the widespread and often sporadic information available on the subject and incorporated it into a more readable and usable format for the reader. Dahlberg also focused greatly on some of the air rage triggers in order to identify prevention strategies. The book contains a wealth of information pertaining to the perpetrator, governing agencies, airlines, law enforcement, and steps the international communities are experimenting with to contain passenger outbreaks pertaining to all phases of flight.

Dahlberg extensively examined many causative factors of air rage. Notably, one increasing risk factor of air rage is brought upon by passenger-felt stress

caused by internal as well as external factors. Internal factors during flight may include medical or neurological factors, alcohol or nicotine withdrawal, or psychological issues. External factors focus around environmental conditions. These issues are more preventable in nature and can include confusing terminal layouts, inaccessibility for disabled citizens, non-courteous crew members, cramped surroundings, and excessive cold or hot temperatures. Some airlines recently have increased external stress factors by reducing seating space per passenger to increase generated revenues. This has caused an infringement on the passenger's personal space and stress is now caused due to the confined cabin environment.

Dahlberg also focused on the increasing restrictions placed on airline passengers. Passengers are restricted in their personal movement and time causing an independent individual to take on a more subservient/dependent role. The higher societal position a passenger holds, the more opposition there is with regards to restricted movement and time. These restrictions are perceived by airline passengers from arrival at the airline terminal to departure of the destination airport. These problems will be intensified with the introduction of the Airbus A380 aircraft, which can be configured to hold 966 passengers in an economy design. It should be noted that Dahlberg discussed passenger misconduct occurring on larger aircraft and longer flights, which is inherent in the new Airbus design.

The author provided multiple misconduct resolution strategies that should be incorporated by the airline industry. These strategies include the improvement of external factors; detecting and assisting with needs; detecting potential passenger misconduct; diffusing potential altercations; educating passengers adequately; and providing choices when available. Many of these strategies require extensive education of crew and airline members. The airlines and national governments also have tried to limit occurrences of passenger misconduct through civil and criminal penalties.

Since the problem of air rage has only recently been brought to the attention of the flying public, only limited research has been accomplished. More research in this area needs to be performed. A definitive testing tool should be developed and utilized to measure and report universal outbreaks of air rage as well as diffused potential air rage incidents. Airlines need to have standard reporting procedures for these particular incidents. Dahlberg explored the airlines' fear of reputation damage and loss of revenue due to reports of air rage; however, she did not delve into air rage between crew to crew members and more research is needed in that area.

This text is a must-read for all aviation employees, including air marshals and educators. The public also would greatly benefit reading this book – to better understand policies regarding passenger misconduct and what is being done to prevent outbreaks. Current human factor courses at the university level should consider utilizing this text, even in a cursorily fashion.

There is no doubt that air rage is a growing problem within the flight industry. Extreme cases have caused procedural deviations of the flight, thus directly affecting the safety of flight. Air rage has been linked to two deaths, which surprisingly, has been the perpetrators of the crime.