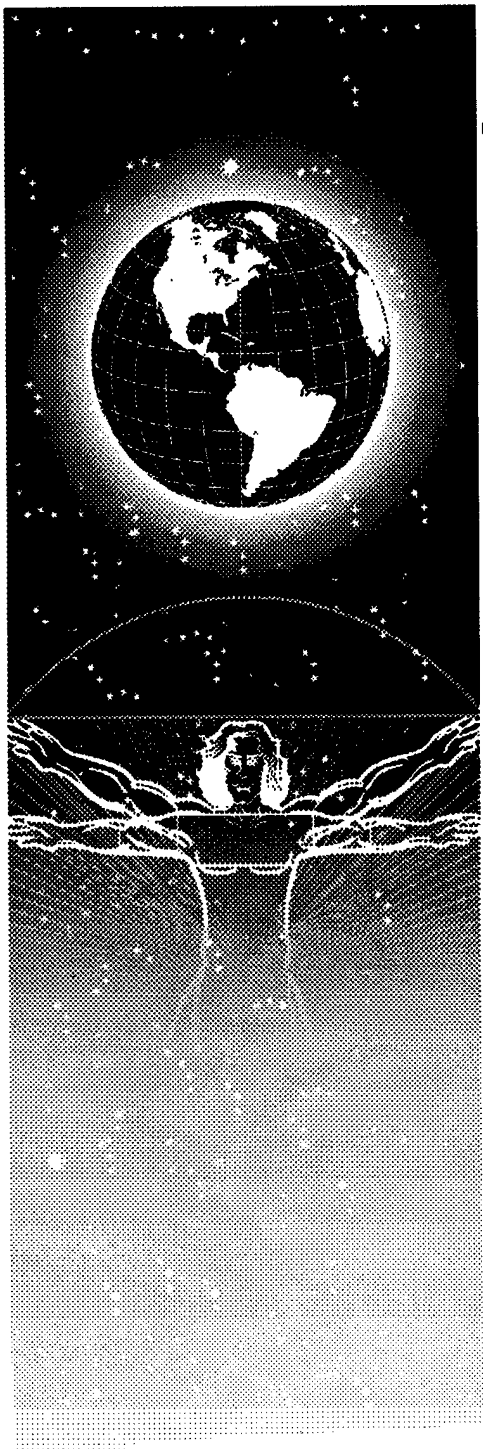
AL/CF-TR-1997-0059



UNITED STATES AIR FORCE ARMSTRONG LABORATORY

PILOT PERFORMANCE VARIABLES

Mark J. Bates
Catherine D. Colwell

UNIVERSITY OF LA VERNE
31-147 OAK STREET
ELMENDORF AFB AK 99506-1400

Raymond E. King

CREW SYSTEMS DIRECTORATE
HUMAN ENGINEERING DIVISION
WRIGHT-PATTERSON AFB OH 45433-7022

Frederick M. Siem
Warren E. Zelenski

AL/HRM
7909 LINDBERGH RD
BROOKS AFB TX 78235-5352

MARCH 1997

INTERIM REPORT FOR THE PERIOD AUGUST 1996 TO MARCH 1997

Approved for public release; distribution is unlimited.

Crew Systems Directorate
Human Engineering Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1997-0059

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

KENNETH R. BOFF, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1997	Interim, August 1996-March 1997

4. TITLE AND SUBTITLE

Pilot Performance Variables

5. FUNDING NUMBERS

PE: 62202F
PR: 7184
TA: 10
WU: 44

6. AUTHOR(S)

*Bates, Mark J., *Colwell, Catherine D., **King, Raymond E.,
***Siem, Frederick M., ***Zelenski, Warren E.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of LaVerne
31-147 Oak St
Elmendorf AFB AK 99506-1400

AL/HRM
7909 Lindbergh
Brooks AFB TX 78235-5352

8. PERFORMING ORGANIZATION

AL/CF-TR-1997-0059

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory, Crew Systems Directorate
Human Engineering Division
Human systems Center
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This review integrates some of the many studies on pilot performance for selection and screening. Pilot performance studies have come a long way as the aviation systems and the roles of the aviator have evolved. No single construct, or operationalization of variables, fully addresses pilot performance. Rather, a multi-disciplinary and multi-modal approach, using significant developments from recent studies, holds the most promise. Developments in information processing theory, at the individual and group levels, combined with findings from safety studies offer potential future criteria of operational performance. In the past, the use of these variables was limited by the difficulty of measuring their impact on performance. Inter-disciplinary efforts, longitudinal studies, and technological advances (such as simulators and computers) enable researchers to better measure these variables. (The combination of previous/current pilot training measures, process-oriented intra/inter-personal variables, safety studies, more functionally and process-oriented personality inventories, and advances in technology/research methods provide opportunities for stronger measures and, ultimately, predictors of pilot performance.)

14. SUBJECT TERMS

Crew Resource Management (CRM), Pilot Selection, Situational Awareness, Stress, Training, Workload

15. NUMBER OF PAGES

110

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89) Prescribed by ANSI Std Z-39-18
298-102 COMPUTER GENERATED

This page intentionally left blank.

TABLE OF CONTENTS

This page intentionally left blank.

CHAPTER I

STATEMENT OF PROBLEM

There are many reasons for studying pilot performance. The major ones include: the expense of training pilots to mission-ready status, the high level of risk and complexity involved in modern aviation, and the threat of potential loss of valuable resources - both equipment and human lives. Therefore, the goal of pilot performance research is "screening out" poor pilot candidates and "selecting in" the individuals who will get the job done well and safely. These issues are of interest to the military, society, aviation regulatory agencies, and aviation industries such as airlines, aircraft manufacturers and training companies. As aviation evolves, these concerns remain at the center of aviation research. Finally, assessment of the optimal amount of automation is an emerging concern.

Military pilot training is a costly endeavor which is made even more costly by attrition during training. The initial training costs of a military pilot are roughly $800,000 based on a 1987 report (Stokes & Kite, 1994), not including subsequent training costs in the operational aircraft and mission. For example, the total training cost of an F-111 pilot is $1.3 million (Driskell & Olmstead, 1989). Even if an Air Force pilot candidate is eliminated during training, the accrued costs are approximately $64,000 in Fiscal 1984 dollars (Bordelon & Kantor, 1986).

The role of human, machine, and environment are becoming more complex and riskier. In general, aircraft are getting faster and more technical. Aircraft flew 100 mph during the First World War. Today, the fastest modern aircraft fly faster than 2,000 mph (Driskell & Olmstead, 1989). Many other innovations have increased the overall complexity including systems sensors and warnings, autopilot, autothrottles, Inertial and Global Positioning Systems, fly-by-wire controls, and mission specific technology such as radar, multiple ordinance, night vision goggles, ground collision avoidance systems, and traffic alert/collision-avoidance systems. Technological changes also reflect the shifting role of the aviator from manual control to flight management (Chambers & Nagel, 1985). In fact, Navy research by Blower (1992) surmises that the next generation of aircraft will place even less importance on manual skills and the pilot will become even more of a manager of complex systems using cooperative human-machine problem solving. The good and bad implications of aviation technology is that the rate of machine failures are decreasing but the number and percent of human errors are rising (Driskell & Olmstead, 1989).

Pilots play a major role in aviation accidents. Even though the Naval Class A mishap rate has decreased from 1953 to present, 58 percent of 308 total Class A mishaps between 1986 and 1990 were due to aircrew error (Yacavone, 1993). Another Naval study between the years of 1972 and 1992 shows the ratio of mishaps attributable to human error versus those due to mechanical/environmental factors going from 1:1 to 9:1 for single-piloted and 12:1 for dual-piloted aircraft (Shappell & Weigmann, 1996). These numbers equate to roughly 150 naval aviation mishaps yearly being caused by human error (Shappell & Weigmann, 1996). Commercial aviation shows similar statistics. According to the National Traffic Safety Board (NTSB) Review of Flightcrew-Involved, Major Accidents of US Air Carriers, 1978 through 1990, the contributing actions or inactions by the aircrew are evident in the majority of fatal air carrier accidents. Overall, human error is estimated to be responsible for 50 to 75 percent of preventable deaths in civilian and military flying (Kohen-Raz, Kohen-Raz, Erel, Davidson, Caine, & Froom, 1994). It is estimated that 80 percent of aircraft accidents are caused by human factors and of these, 80 percent are related to disorientation and loss of situational awareness (Popplow, 1994).

Aviation Environments

There is a wide variety of aviation environments where pilot performance can be measured. There are some elements that are common to all aviation settings and some that are unique to a particular airframe. The airframes can vary in size, number of aircrew, type of propulsion (e.g. jet, propeller, helicopter), mission goals (e.g. training, transport, aerial refueling, search and rescue, bombing, air combat), mission hazards and stressors (e.g. instructing students, low levels, transiting foreign countries and airfields, long and irregular crew duty days, crossing many time zones, inclement weather, high traffic, combat), technological equipment, and organizational structure (e.g. Air Mobility Command, Air Combat Command, Special Operations, major airline, commuter service, aeroclub). Analyzing the demands and characteristics of the different airframes would benefit pilot performance research but is beyond the scope of this review. This review concentrates on the factors common to the performance of transport pilots, assuming that the majority of these factors are applicable to all pilots.

Aviation Stresses, Tasks, and Goals

The most important aspects of a pilot's job are illustrated by the stresses, tasks, and goals of flying. The stresses and tasks of flying include timeliness, speed, multiple sources of information, coordination responsibilities, multitasks, and prioritization. Ironically, there is also a threat of understimulation and inadequate feedback due to technological automation. This combination of stresses and tasks creates the challenge and risk pilots enjoy. Moreover, because of the difficulty and danger involved, the goals of flying are accomplishing the mission effectively and safely.

In both military and commercial aviation, there is normally pressure for timeliness. Commercial aviation is concerned with timeliness, coordinating many departing and arriving aircraft and providing on-time service to passengers. Likewise, timeliness can be crucial for military operations coordinating efforts with other aircraft and forces. Not being timely in military operations can mean losing the tactical advantage, not backing up other forces, and ultimately losing lives and equipment.

Not only is there an emphasis on being timely, aviation is inherently fast and there is no opportunity to "stop on the side of the road." The critical phases of flight, takeoffs and landings, last only several minutes yet there is an enormous amount of information to monitor and only seconds for decision making. Flight Safety International instructors refer to these critical phases as "safety windows" or "windows of risks" because the majority of accidents occur during takeoff/climbout and approach/landing. According to a NTSB (1994) safety study on flightcrew-involved major accidents of US air carriers, 27% of accidents were during takeoff and 51% during landing. This area or box is often arbitrarily assigned 2,000 feet and below where the potential risk increases the closer the aircraft is to the ground.

The speed of flying also refers to rate and number of tasks and information sources. There are combinations of related and unrelated tasks that must be accomplished during certain phases of a flying mission. Attention and energy must be efficiently spent. For example, during takeoff and landing, most of the attention is focused on monitoring and controlling the aircraft's flight path, speed, and configuration. Besides actually flying the aircraft, there are tasks such as monitoring instruments, coordinating on the interphone and radios, and running checklists. At any time during a mission, pilots may juggle coordination with other crew members in their aircraft, crew in other aircraft, Air Traffic Control, weather, and an operations center. Moreover, all of the aircraft's systems need to be continually monitored by checking gauges and indications. Any kind of abnormal indication needs to be taken care of as well as analyzing implications to other systems and the mission. For example, there are 455 separate warnings and caution alerts on the Boeing 747 with minimal built-in prioritization or filtering of these indications (Chambers & Nagel, 1985). All of this equates to doing several things at once and prioritizing cues and tasks. When the stresses and tasks are assembled, there is a dangerous potential for accidents due to overstimulation and too much information.

There are also dangers associated with understimulation and inadequate feedback due to the increased automation in flying. The pilot may become bored and complacent as the passive monitor of the plane's automated systems and relinquish too many responsibilities to automation features. Relying on automation can create situations where pilots assume everything is taken care of and become less vigilant (Foushee, 1982). The problem is that automation technology is like a third crew member who flies and follows directions well but has no common sense. The automated system will do what it is told even if that means descending into terrain or transferring all of the fuel out of one tank and flaming out an engine. The pilot may delegate too much to the automated systems and not ensure that the task is carried out correctly. Detection of these interaction errors with automatic systems is even more difficult because of possible long delays before a problem becomes apparent (Chambers & Nagel, 1985). As aircraft become more technical, the interface difficulties between man and machine will become more important.

The pilot's primary goals are doing the job well and safely. Most of aviation research can be broken down into studies about either performance or safety, but rarely both subjects together. Actually, these two aspects of flying are highly related. A job can be done well (e.g. accomplishing the desired objective) but not safely (taking too many risks), safely but not well (not taking reasonable risks to accomplish the mission in an effective or timely manner), or not well and not safely. Often there is a tradeoff between operational effectiveness and safety precautions that forms a continuum. This continuum is represented by an often-present tension between operations and safety agencies. Likewise, Alkov (1987) shows that when the balance shifts toward increased operational demands in response to international crises, safety considerations may be compromised and accident rates increase. Along these lines of reasoning, any evaluation of performance should take into account safety measures, whether selecting positive behaviors or screening out undesirable behaviors.

The importance of safety is underscored by military and federal flying regulations and checkride procedures. A primary concern of flying regulations is ensuring safety of flight. Some of the safety guidance includes minimum weather, fuel, equipment, crew complement, crew rest, training currency and instrumentation requirements as well as maximum crew duty times. The same emphasis on safety issues is evident in military and civilian checkrides. For example, the majority USAF checkride items can be graded marginal except for three: Emergency Procedures, Safety, and Judgment. These items do not represent accomplishing maneuvers accurately or expeditiously, but rather doing maneuvers safely within military, aircraft, and personal limitations.

Pilot Attributes

The basic attributes of an aviator are aeronautical knowledge, skill, and judgment (Buch, 1984). Situational awareness and attention to detail are often considered essential to good judgment and are also important attributes. As a case in point, military instructor pilots frequently use the catchwords "attention to detail," "situational awareness," and "judgment" when debriefing a flight. Knowledge and skill are the tools and judgment reflects the fundamental process of how the tools are used. These attributes increase in breadth and depth as a crew member matures as an aviator.

CHAPTER II

STATEMENT OF PURPOSE

No single construct or operationalization of variables fully addresses pilot performance. Rather, a multi-disciplinary and multi-modal approach, using significant developments from diverse studies, has the most promise. Pilot performance studies have come a long way just as the aviation systems and the roles of the aviator have evolved. There are numerous isolated developments using pilots' perceptions of performance, safety studies, and more-appropriate personality measures as well as more sensitive and broader ranging measurements and technology. In addition, crew resource management and human factors concepts offer systemic models of performance. These systemic models, constituting personal and interpersonal resource management, measure the essential dynamics that can be misrepresented in outcome studies. Systemic models offer the framework to potentially integrate other developments and meaningfully explain the complex interaction of cognitive, affective, and behavioral patterns with the time-critical multi-task and multi-resource aspects of pilot performance.

While the aims of aviation psychology have not changed, the methods have. In the past, the use of these variables was limited by the difficulty of measuring their impact on performance. Early aviation research used psychomotor equilibrium tests on a device that resembled a ski lift chair in which the seated individual did acrobatics (Koonce, 1984). There was also an emotional stability test measuring changes in pulse rate and respiration when a pistol was fired behind the pilot candidate's back (Koonce, 1984). As aviation psychology develops, research uses broader-reaching variables and finer measurements. Today, inter-disciplinary efforts, advanced methods of testing and technological advances, such as simulators and computers, enable researchers to better measure these variables. The combination of present outcome-based pilot training measures and potential process-oriented variables provide opportunities for stronger measures and predictors of pilot performance. A systemic perspective can integrate many different variables to offer a functional and holistic perspective on measuring and predicting performance.

This review outlines and integrates some of the many studies conceptualizing and measuring pilot performance. The performance studies are organized by variable category including pilot training, safety studies, pilot perceptions of performance, cockpit resource management, and human factors. Variable definitions and general characteristics are introduced and then various measures are reviewed in context of the instruments, methods, and results. After the review, an integration of the variables is proposed and illustrated with potential criterion measures, predictor variables, and research directions.

CHAPTER III

PILOT TRAINING VARIABLES

Pilot Training Criterion Measures

For the purposes of this review, pilot performance is the criterion variable. The pilot's desired performance can be measured as completing a mission or flying program, completing tasks that make up a mission, or how the pilot accomplishes these tasks during flight. In other words, performance can be looked at as a whole event, a select number of critical tasks, or the critical processes that are essential to complete the flying tasks. The majority of pilot performance studies focus on Undergraduate Pilot Training (UPT) success to validate screening and selection methods. However, UPT performance does not necessarily equate to operational performance, which is the true objective of all screening and selection.

The most frequent and valid criterion measure is UPT performance despite several limitations. Even though there are many pilot performance studies, finding suitable criterion measures of performance is difficult. Pilot training success remains the primary measure because it is the last opportunity to measure success under uniform, controlled circumstances with easily available results. These measures include passing UPT, checkride grades, flight grades average, time to complete UPT, and airplane assignment after UPT (Fighter-Attack-Reconnaissance qualified versus Tanker-Transport-Bomber tracks). Each of these measures have logical and statistical limitations in determining which pilots succeed operationally.

A major logical concern is that pilot training performance is not necessarily the same as operational performance. Research based on pilot training success suggests who is most likely to succeed in pilot training, but may have limited correlation with who does well operationally. Bale, Rickus, and Ambler's (1973) study on naval aviator progression shows a decreasing predictive power the further an aviator progresses in the flying career. This study suggests that pilot training does not measure all "mission- oriented" abilities.

Pilot performance may also change as the result of going from a training to operational environment. Helmreich (1986) identified a "honeymoon effect" on motivation that may lead to inflated expectations of operational performance. This effect shows that high levels of motivation may be associated with training environments or initial stages of employment and subsequently lag during the real job, resulting in poorer performance.

Checkrides, daily flight grades, time to complete training, and even aircraft assignment can be affected by subjective elements. The different personalities and philosophies of instructors and evaluators are inevitably part of the training and evaluating process. The content and delivery of a checkride can vary according to what the examiner thinks is important. There is no evidence of consistent grading criteria. In a study on naval instructors' flight evaluations, Dolgin, Gibb, Nontasak, and Helm (1987) could not identify strong clusters of training items that accounted for the majority of instructor grading variance. Because instructors represent the greatest source of grading variability, McDaniel and Rankin (1991) have proposed using a mathematical decision aid to improve the reliability and accuracy of instructor grading.

In addition, there is normally a primary instructor who may affect the student by high or low grading standards or a personality conflict or attraction. Using the Myers-Briggs Type Indicator measures, Kreienkamp and Luessenheide (1985) showed that the amount of time the student needs to learn was to be significantly affected by differences in personality with a given instructor. Likewise, a positive or negative 'halo effect' may influence an instructor's expectations and perceptions of performance (Stokes & Kite, 1994).

Research studies sometime use fighter-attack-reconnaissance (FAR) assignments as the criterion of top performance. Top pilots, however, do not always select, nor are they always given, FAR assignments. The eventual aircraft qualification or assignment can be influenced by subjective factors of both raters and students. The process of ranking top pilots is partly dependent on the subjective judgment of the raters. This is further confounded by the top pilots who do not want fighter-type assignments. There is potential for student pilots to make their preferences known and influence aircraft assignments through varying their own performance and influencing the instructors who determine aircraft assignments. Simply what's available at the time of graduation is often a major assignment driver.

There are several other variables that are not ability-related but may confound UPT performance measures. Student pilots may fail or withdraw for medical or motivational or academic reasons as well as inability (Gibb, 1990). Likewise, the time to complete UPT can vary according to medical and administrative breaks in training.

Organizational demands may also arbitrarily affect pilot training and even operational performance measures. The military may change the desired success rate, artificially affecting how many pilots pass and fail (Damos, ****inpress). Even operational upgrades may be more dependent on unit manning requirements than when a pilot is ready for upgrade.

There are also inherent statistical limitations imposed by the narrow range of pilot training measures. Daily flight grades, checkride scores, and especially pass/fail rates have limited variability. The dichotomous pass/fail variable imposes a range restriction on variance between students that underpredicts relationships (Gibb, 1990; Hunter & Schmidt, 1992; Jackson & Ree, 1990). There is also further range restriction because of limited attrition. Likewise, checkride failures are also rare which limits the ability to discriminate effectively between pilots. Even the informal and formal screening processes of competing for UPT entry impose a range restriction as certain groups of applicants are eliminated. Some of these screening processes include passing college, physicals, intelligence- (or *g*-loaded) tests, officer training programs, and the Flight Screening Program (FSP). For example, Stoker, Hunter, Kantor, Quebe, and Siem (1987) demonstrate FSP reduces pilot training attrition (although, reduced UPT attrition is apparently due more to an experience and training effect than screening success). Ironically, early-program screening and training statistically limit the ability to discriminate between other predictors.

There are some additional factors that might confound using measures of pilot training success as criteria. Previous flight time and age have been found to predict flight performance (Gibb, 1990). In fact, previous flight experience has been shown to add the greatest incremental validity beyond the Air Force Officer Qualifying Test (AFOQT) composite compared with variables such as psychomotor and information processing measures (Carretta & Ree, 1994). These factors probably only affect the initial success of most student pilots, yet grade averages would reflect an early advantage.

An extreme example of the effects of previous flight experience is the former navigators who crosstrain. These student pilots have already proven themselves in the military aviation environment to be selected for UPT. Prior navigators enjoy the advantage of both prior military and operational experience compared with the average UPT student who just graduated college and completed flight screening in a single-engine propeller plane.

Pilot Training Predictor Variables

The predictor variables are factors that should explain pilot performance variance. There are five general categories of predictor variables of pilot performance. According to Street, Helton, and Dolgin (1992b), the most to least robust predictor variables are: 1) psychomotor coordination, 2) background information, 3) information processing ability, 4) general cognitive ability, and 5) personality traits. Each variable is explained by a general definition, different measures and methods associated with it, and practical considerations (the ease and limitations of administering and taking the tests, including time and costs involved, and the reliability and validity of the results).

Whenever possible, these predictors are compared with the most common and available validity yardstick, pilot training success. Therefore, it is important to keep in mind the limitations of using pilot training success as the ultimate criterion. Likewise, research may be reaching a ceiling on how much additional UPT variance can be explained. This would be consistent with a meta-analysis on predicting pilot-training success which shows a trend of decreasing validity coefficients over the years (Hunter & Burke, 1994). Therefore, the small incremental predictive validities eked out by additional predictors may be a function of efficiency of the established measures.

General Cognitive Ability

The general cognitive domain has been the most widely tested domain in actual pilot selection (Street et al., 1992b). Several studies show general cognitive ability, psychometric *g*, is important in the prediction of job performance (Hunter & Hunter, 1984; Olea & Ree, 1994). In addition, tests of general cognitive ability are low cost and easily administrated to large groups (Street et al., 1992b).

The tendency of pilots to have superior general intelligence is well documented. Studies of UPT students find the average IQ to be around 120, more than one standard deviation above average and in the high average range (King & Flynn, 1995; Retzlaff & Gibertini, 1988). Similarly, a study of Air National Guard F16 pilots' Multi-dimensional Aptitude Battery (MAB) scores demonstrated superior intellectual functioning (Flynn, Sipes, Grosenbach, & Ellsworth, 1994).

All military services employ their own version of a largely cognitive screening inventory. Until 1993, the US Navy and Marine Corps used a paper-and-pencil Academic Qualification Test (AQT) measuring flight-related academic abilities and the Flight Aptitude Rating (FAR) measuring aptitudes. In 1993, the AQT/FAR Aviation Selection Test Battery was revised to the ATSB. The US Army (USA) use similar paper-and-pencil tests called the Flight Aptitude Selection Test (FAST). The Air Force (USAF) uses the Officer Qualifications Test (AFOQT).

The AFOQT is a good example of how these cognitive tests are designed. The AFOQT is composed of 16 tests, including 3 power tests, 3 primarily speeded tests, with the remainder being mixed power and speed tests (Skinner & Ree, 1987). The tests are assembled into five classification composites: Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Navigator-Technical (N-T). The composites measure differential aptitude and are all highly *g* saturated (Olea & Ree, 1994).

The Federal Aviation Administration (FAA) has also proposed using mental status exams as part of the aviation medical exam. However, Banich, Stokes, and Elledge (1989; Stokes, Banich, and Elledge, 1991) all conclude that such a mental exam focuses on too low of a level of cognitive ability and would not measure some cognitive skills that are required to be a pilot. Using clinically normed tests on a high functioning, non-clinical population is also problematic. Instead, studies recommend that any kind of clinical assessment of cognitive functioning focus on cognitive abilities fundamental to pilot tasks such as: 1) perceptual-motor abilities; 2) visio-spatial abilities; 3) working memory; 4) attentional performance; 5) processing flexibility; 6) planning or sequencing abilities; and 7) risk evaluation (Banich et al., 1989;

Stokes et al., 1991). These specific cognitive abilities fall under the psychomotor and information-processing categories of dependent variables.

Nevertheless, specific cognitive measures have been found to provide little additional predictive ability beyond general intelligence measures. Olea and Ree's (1994) pilot training study found little difference between the predictive efficiency of specific ability or job knowledge (*s*) and general cognitive abilities (psychometric *g*). These researchers conclude that general cognitive ability is the best overall predictor of job and training performance (Olea & Ree, 1994). Another study comparing the results from a multiple aptitude cognitive test and psychomotor battery found high average multiple correlations implying that psychomotor tests could provide only small additions in validity to cognitive measures (Ree & Carretta, 1994).

Psychomotor Coordination

According to Street, Helton, and Dolgin (1992b), psychomotor strategies "typically focus on eye-hand-foot coordination in their simplest forms, although more promising strategies have combined such skills with information processing, problem solving, and reaction time in an aircraft-like environment" (p. 1). Perceptual-motor abilities include precisely controlled and coordinated movements of two or more limbs in response to dynamic stimuli (Tirre & Raouf, 1994). Basically, the psychomotor tests tap into the mental and physical coordination of thinking about and doing several tasks at once or sequencing multiple tasks in a short period of time. There have been several approaches to measuring psychomotor abilities, mostly utilizing computer technology with controls and tasks that simulate flight.

Cox (1988) describes the development of the Air Force's psychomotor measures. The Air Force started with electromechanical versions of Two Hand Coordination (2HC) and Complex Coordination (CC) psychomotor tests which evolved, with advances in computer technology, into a psychomotor and information-processing test battery as part of a more comprehensive Basic Attributes Test (BAT). The test is conducted on a portable testing station that utilizes a joy stick controlling the horizontal and vertical movement of a piper on a monitor.

Research supports the reliability and validity of the Air Force Two Hand and Complex Coordination Tests. This psychomotor data has been shown to account for 5-11% of the variability of UPT outcome depending on how performance is measured (Cox, 1988). When hours required to complete UPT were added as additional criteria, the multiple correlation with psychomotor measures was 0.52 compared to 0.56 with the UPT outcome (pass/fail) criterion (Cox, 1989). The author concludes that the high results were probably due to a sampling artifact and there was no significant difference between the criteria of UPT outcome and UPT hours. In another study, Bordelon and Kantor (1986) showed that measures of psychomotor ability differentiated candidates likely to graduate UPT as well as to receive superior ratings (FAR-recommended).

The Navy has done similar research with computer-based psychomotor tests (CBPTs). The Psychomotor Test/Dichotic Listening Test (PMT/DLT) is the most powerful CBPT (Delaney, 1992; Street & Dolgin, 1993). Seven subtests are used to measure eye-hand-foot coordination, divided attention, and selective attention. The PMT monitors simulated stick, rudder, and throttle control movements as subjects move cursors on a computer. The DLT measures differences in selective attention to different digit and letter sequences presented to each ear simultaneously. Using a criterion of flight grades, Delany's (1992) study found a high correlation with psychomotor scores and a moderate correlation with dichotic listening scores. In the same study, the PMT performance accounted for 19.5% flight grade variance which was largely independent of the 16.6% variance described by the current selection tests and demographic variables. Another Navy study demonstrated that performance on certain psychomotor tests could make a modest improvement in training assignments (between jets and other aircraft assignment "pipelines" requiring lower performance standards) and training performance (Street & Dolgin, 1993).

Gopher's (1982) study with Israeli flight students also shows the effectiveness of dichotic listening tasks (measures of selective attention) in predicting student pilot success. Dichotic tasks send different information simultaneously to each ear. Three types of selective attention measures, omissions, intrusions, and switching errors, had high correlations between themselves and low correlations with other pilot selection measures. Although the group of students who completed training made less errors on all three types of selective attention than the students who did not complete training, the addition of attention measures made a relatively small contribution to the selection battery.

Boer and Castelijns (1991) has used the Processing Information under Loading and Time sharing conditions (PILOT) to successfully test pilot applicants in the Netherlands. There were modest but significant correlations between the psychomotor test scores and flight grades. The PILOT test was also compared with the Precise Coordination Multitask Process (PCMP), another test for psychomotor speed. The two psychomotor tests both measured "tracking tasks" and were correlated (.46), but the PCMP had no validity indicating that differences in test constructions may have significant effects on validity.

Many of the psychomotor tests employ computer-game type apparatus which can also be used for training. Gopher, Weil, and Bareket (1994) conducted a study on the transfer of skills from a complex computer game to flight performance suggesting that computer games can improve and generalize psychomotor skills to new situations (Gopher, Weil, & Bareket, 1994). Israeli flight cadets who were trained in a computer game (the training focused on specific skills involved in playing the game) were compared with cadets who played the game without instruction (who were expected to gain an ability to cope with high processing and response demands and learn better attention control) and cadets who did not play the game. Both game groups performed significantly better than the no-game group.

Since the various psychomotor tests rely on computer-generated images and controls, there is concern about the possible confounding influence of previous experience on similar computer tasks like video games. Tirre and Raouf (1994) found home or arcade video game performance benefiting men's but not women's flight simulator performance. These differences in gender performance based on video-game experience may be due to how a gender's earlier experiences and preferences might transfer differentially to a given type of task (Tirre & Raouf, 1994). An expected result of the same study was the correlation between higher general cognitive ability (*g*) and the psychomotor performance.

Psychomotor measures often correlate highly with information processing measures and both measures can be subject to potential confounding effects of video-game experience as well as high correlations with *g*. Both psychomotor and information processing measures target more complex, multi-task abilities that are often best measured on a computer apparatus. Therefore, both psychomotor and information processing may be subject to previous video game experience affects because of similar types of manipulating tasks where practice changes performance because it is not a novel task anymore. Likewise, both measures simulate higher functions of general cognitive ability and consequently, are likely to duplicate explanation of variance with cognitive ability. Tests with US Air Force recruits show a high correlation between *g*-saturated Armed Services Vocational Aptitude Battery (ASVAB) and the BAT psychomotor tests (Ree & Carretta, 1992). The average fully corrected correlation was .73 implying that the psychomotor tests were also *g* loaded. This indicates that psychomotor tests may provide little incremental validity beyond the standard *g*-loaded screening tests.

Biographical Information

Biographical Information reflects what a person has done in the past. Background tests are generally thought of as the best predictor of early naval training attrition (Hilton & Dolgin, 1991). The background information should present a stable representation of a person's interests and attitudes (Street & Dolgin, 1992a). The assumption is that a person's prior knowledge and interest in aviation predicts future interest in an aviation career. This measure potentially taps into important motivational factors that are otherwise not measurable. Although, since the background test is susceptible to self-report bias, questions should focus on real-life situations or actual experiences. Background testing can provide valuable information when conducted well and with select groups.

Street and Dolgin (1992a) found statistical differences in many Aviation Selection Test Battery (ASTB) Biographical Inventory (BI) responses of naval student pilots passing and failing preflight training. The results indicate that the ASTB BI could be used to reduce the number of aviation cadet attrition by 50% at a cost of non-selecting approximately 20% of those who would have succeeded (Street, 1992). The Biographical Inventory would be used to provide additional screening for the candidates from other-than ROTC and service academy sources. This inforamtion would benefit both candidates and the Navy because of the relatively high attrition rate of those candidates in preflight training.

The potential value of biographical screening may be limited to candidates who were not already rigorously screened through ROTC or by a service academy and only in the early stages of flight training. One of the difficulties of studying pilot performance variables is the inherent range restriction imposed by flight physical standards, college education, and ROTC or service academy selection and military education. Pilot candidates are already a unique group. Any group of pilot candidates who has not undergone a stiff screening process or chosen a military education might benefit by biographical tests.

The branches of the military, NATO countries, and civilian aviation place differing emphases on biographical data. The Navy FAR/AQT changed to an aviation questionnaire with more biographical information. USAF flight surgeons conduct a semi-structured interview, termed the Adaptability Rating for Military Aeronautics (or Aviation, ARMA), to assess motivation to fly and solicit limited biographical screening during the initial flight physical. The flight surgeons ask questions about "aviation affinity" (why candidates want to be a pilot) as a brief and crude way to detect unsuitable candidates (Mills & Jones, 1984). The ARMA, however, is inconsistently used and flight surgeons are not satisfied with it (Verdone, Sipes, & Miles, 1993). NATO countries rely heavily on biographical data in pilot selection to provide a complete individual profile (Street & Dolgin, 1992a). Civilian aviation requires biographical data, but with a different candidate population who are already proven pilots.

Information Processing

Information processing "concerns how people attend to, select, and internalize information and how they later use it to make decisions and guide their behavior" (Corsini, 1994, p. 245). The present information processing variables make little contribution to predicting performance. Nevertheless, the information processing strategies appear to be an important factor in how pilots cope with the complex aviation tasks and environments so research continues to look for meaningful relationships with performance.

Information processing concepts are used in several military screening tests and research studies. Common instruments include the US Air Force's Mental Rotation and Item Recognition from the BAT and the US Navy's Complex Visual Task. The Mental Rotation is a spatial transformation task determining whether pairs of letters are the same or mirror images and in the same relative position or rotated in relation to each other (Carretta & Ree, 1994). The Item Recognition is a measure of short-term memory presenting one to six numbers on a screen and then presenting a single number, asking if the single number was one of the original numbers. Using UPT pass/fail and class rank criteria, Mental Rotation and Item Recognition yielded an incremental validity, .006 for both criteria, compared with the incremental validity for psychomotor predictors of .039 for pass/fail and .038 for class rank.

Fowler (1981) compared The Aircraft Landing (AL) test measures of information processing with two flight training test scores of Canadian Forces student pilots. The AL test measures hierarchical mechanisms and feedback as two dimensions of information processing. The hierarchical mechanisms attempted to measure lower level processes (such as attentional selectivity) as well as higher level processes (such as learning) that adapt to environmental demands by increasing its effective channel capacity. Likewise, feedback measures investigated individual differences in utilizing feedback effectively, a critical feature in the learning process. In this experiment, two groups of Canadian student pilots, with and without previous flying experience, were evaluated. The progress of these pilots was evaluated until reaching a criteria skill level in a device where simulated approach and landings could be attempted and learned. The test scores showed validities up to 0.49 against the criterion of flying tests in light aircraft at the 7 and 12 hour point in training. An information processing model of skilled performance was chosen over an abilities classification model. The study suggests that monitoring the time in trials to mastery while teaching a new and complex task may be a valuable way to measure the learning dimension of information processing.

Fedor, Rensvold, and Adams (1992) investigated the information process of seeking feedback with helicopter pilot trainees. Feedback can be sought by asking directly (eliciting) or using indirect means such as observation (monitoring). These two ways of getting feedback will yield different types and amounts of information. Both of the examined factors, individual differences and situational variables, were significant predictors of different feedback-seeking behaviors. Although the study did not look for influences on performance, the number of factors and interactions illustrates the complexity of measuring and understanding process variables. Significant interactions were found between tolerance for ambiguity and feedback seeking costs as well as self-esteem and source credibility.

Different periods of training may differentiate between the relative importance of psychomotor and information processing measures as well as limit their contribution to predictive validity. Boer and Castelijns (1991) describe English naval aviation training reports that early training failures were due more to mechanical skills where later failures were due more to time-sharing (conceptualizing time-sharing skills as the ability to do several tasks at once such as monitoring a source of information while doing an independent task). Likewise, Damos (1993) proposes that multiple-tasks may represent a time-sharing skill that is more important in advanced stages of training or when there is limited time to complete a task. Therefore, single-task measures could have more relevance with early pilot training performance and dual-task measures with later pilot training. By the time complex information processing may become a factor in determining performance, additional screening built into the academic and flying portions of earlier training impose a further range restriction.

Personality

Day and Silverman (1989) suggest that occupational selection strategies might benefit by considering personality dimensions that are relevant to the specific job and organization. There have been many efforts to find personality factors that can predict differences in pilot performance. These efforts use a wide variety of approaches and have mixed results. Research is driven by, complicated by, and possibly even enhanced by the wide variety of these approaches. The mixed results of pilot personality research is understood by first comparing the different testing approaches and then reviewing the results of several pertinent studies.

The approaches vary on underlying constructs, instruments, and methods of measurement. Often, researchers disagree on a general model of personality or the number of factors needed to adequately describe human behavior (Digman, 1990). In turn, these constructs primarily drive the choice of what testing instrument and method is used. Researchers also consider the subject population's makeup and environment when choosing the most appropriate, effective, and practical instruments and methods available. There are multiple instruments that are designed for broad or specific personality traits and populations. These instruments fall under the three methods of personality testing: self-report, peer-report, and professional evaluation. Each of these methods have unique and shared advantages and limitations which impact the research results.

Self-report inventories are the most widely used because of their standardization and ease of use. On the other hand, some of these instruments may have low reliability and validity (Retzlaff & Gibertini, 1988). Other limitations are: 1) the susceptibility to faking, 2) the pilot's perception about the function of the test, and 3) the clinical orientation which tends to over-pathologize clinical symptoms and under-discriminate higher functioning qualities. Pilots may try to intentionally or unintentionally make themselves look good on personality tests and, at a minimum, show defensive presentations. Butcher's (1994) study using the MMPI-2 with airline pilot applicants indicated tendencies to minimize adjustment problems and be defensive in attempts to create a favorable impression (although the study did find that the MMPI-2 was better than the MMPI for accurately portraying non-clinical subjects). To avoid this kind of response bias, personality assessments try to mask the dimension of interest from the subject (Street & Helton, 1992b). Another possibility for getting more accurate information is anonymous testing and peer ratings which may allow pilots to disclose more sensitive information about themselves and their peers (Flynn et al., 1994).

Measuring response latency may offer a way to enhance self-inventories (Siem, 1996). Siem explains that differences in response times to personality items can be interpreted as the degree of endorsement or rejection, reflecting the examinee's self-concept or self-schema. In Siem's study, 509 UPT students were asked to answer as quickly as possible the Automated Aircrew Personality Profiler made up of 202 relevant items from different personality inventories. The five scales are labeled as socially desirable characteristics including: 1) Communality/frequency (opposite of Psychoticism/infrequency), 2) Emotional stability (opposite of Neuroticism), 3) Extraversion, 4) Competency (opposite of Inadequacy), and 5) Trusting (opposite of Cynicism). The scale scores and response latencies showed some correlation, but not consistently across all trait dimensions. Only the response latency for the endorsed extroversion scores added incremental validity over the scale scores in predicting UPT graduation. Moreover, the latency-based self-schema scores were less reliable than the associated scale scores.

Another personality measure, not commonly used in pilot studies, is peer reports. Peer reports ask someone who knows the subject on a regular basis to report the subject's normal behavior. The advantages of peer reports are the familiar perspective on a person's personality which avoids self-report bias and describes, it is hoped, consistent behavioral problems that might not otherwise be evident. Nonetheless, peer reports may reflect a different kind of unintentional bias of the peer's subjective interpretation of behavior. For this reason, peer reports should be structured inventories asking specific behavioral questions to avoid as much subjectivity as possible. In addition, due to the highly competitive UPT environment, peers may even intentionally bias reports by refusing to give each other low ratings ("cooperate and graduate") or otherwise misrepresent each other for competitive reasons.

In contrast to peer evaluations are structured interviews and other forms of direct evaluation. The professional evaluation offers the advantages of being done by trained personnel who can perceive more than a self-report or peer. However, the subjective nature of the evaluation still precludes standardization. Furthermore, the structured interview for general pilots is the most time-intensive way to measure personality, especially if only a few professionals are assessing many pilots. In the German Air Force or Luftwaffe, each pilot training applicant is given a diagnostic interview covering stress levels during testing, coping strategies, achievement, flying motivation, and personality characteristics and traits that could affect a career in military aviation (Gnan, Flynn, & King, 1995). In the USAF, the ARMA, as previously noted, assesses healthy motivations to fly. In the case of returning a pilot to flying duty, Adams and Jones (1987) propose that a professional interview is the best way to assess very subtle factors with a typically healthy, well-defended population. Adams and Jones (1987) explain that grounded flyers are usually intelligent, articulate and eager to resume flying duties while also being "rarely attuned or introspective , making them particularly vulnerable to the psychosomatic manifestations of anxiety" (p. 350). The USAF example suggests that there may be select cases where a professional structured interview gives a needed closer look at an individual's psychological functioning.

Another tool for professional evaluations is the projective test, although it is rarely used in the United States. Projective tests, exemplified by the Rorschach ink blot, use ambiguous visual stimuli to elicit responses that are supposed to represent character "projections" (Turnbull, 1992) In Sweden, the projective Defense Mechanism Test (DMT) has been used to identify accident-prone individuals but the test's validity is not established (Turnbull, 1992). The DMT uses pictures shown at a speed below the threshold of awareness and records the incorrect perceptions. These misperceptions are supposed to provide information about the personality structure's defense mechanisms. The assumption is a pilot who invests energy in strong defenses has less energy available for recognizing threats and coping with heavy workloads (Turnbull, 1992).

It is generally difficult to test the personality of the pilot population because of the pilot's guarded attitude towards testing, the pilot's non-introspective personality, and the limited availability of operational pilot samples for norming purposes. Due to the frequent monitoring of both civilian and military pilots' physical and mental health, pilots avoid anything that would put their medical and psychological qualification at risk. In addition, personality testing results may be limited because pilots are speculated to be less psychologically introspective (Picano, 1990; Reinhardt, 1970) and reluctant to admit perceived weaknesses (Flynn et al., 1994). Even if pilots were a cooperative population, it is also difficult to assemble a sample of operational pilots for standardized testing. The research subjects are usually unique samples of volunteers, subjects seeking waivers, or special groups where medical and psychological evaluations are required (Flynn et al., 1994).

The personality research that has been done with pilots has provided mixed results. There is irregular evidence of personality traits predicting performance under different circumstances that is rarely repeated by other studies (Carretta & Siem, 1988; Chidester, Kanki, Foushee, Dickinson, & Bowles, 1990; Gnan, Flynn, and King, 1995; Siem, Carretta, and Mercatante, 1988; Siem, 1992; Street, Helton, & Dolgin, 1992b). However, pilot personality studies are successful in finding significant patterns for the pilot population as well as differences between groups of pilots. Although these studies cannot establish the reasons for these patterns and differences, Chidester, Kanki, Foushee, Dickinson, and Bowles (1990) suggest using the cumulative personality measures to create pilot norms which can illustrate differences from the general population. These norms can also be used to cluster pilots in groups for comparison purposes (Chidester et al., 1990) as well as identify the characteristics associated with a specific aircraft type (Flynn et al., 1994). Moreover, there is speculation that pilots with certain personalities may be more functional in, or attracted to, different aircraft and missions (Retzlaff & Gibertini, 1987).

There are various general considerations for looking at personalities and their measures. One important discrimination is whether a personality characteristic is more permanent or flexible in nature. For example, a series of studies (Chidester, Helmreich, Gregorich, and Geis, 1991; Gregorich, Helmreich, and Wilhelm, 1990; Helmreich, 1984; Helmreich, Wilhelm, Gregorich, and Chidester, 1990; Helmreich and Wilhelm, 1991) concentrated on measuring attitudes which are presumably more flexible and susceptible to training influences. On the other hand, there are many adult personality dimensions that are relatively stable, are not subject to much change, and should be considered more in "screening out" and "selecting in" pilot candidates. Another way to differentiate personality profiles is based on three factors: elevation (what direction and magnitude), scatter (how homogeneous), and shape (how interrelated) (Gregorich, et al., 1990).

In general, the pilot personality has been found to be psychologically stable and adaptive (Butcher, 1994; King, 1994; Flynn et al., 1994). There is a long list of personality traits that are theoretically attractive but do not yield much predictive validity. Nevertheless, there is considerable agreement in what experts and researchers think the personality characteristics of a pilot are.

There are pilot personality models, based mostly on expert observation, that are used for training and research formulation. Hughes refers to one such theory of pilot personality in their unpublished crew resource management workbook. This theory, proposed by naval flight surgeon Frank Dulley, describes four lifestyle characteristics that contribute to success and five potential defects. The four lifestyle characteristics are: 1) Need to be in control; 2) Emotionally-distant opposite-sex; 3) Mission-oriented, compartmentalizing approach; and 4) Systematic and methodical. The five possible defects are: 1) Limited spontaneity; 2) Complacency; 3) "Familiarity-breeds-contempt" syndrome; 4) Ritual trap; and 5) Needing "positive-maleness" ego feedback. Dulley's description, however, is not based on empirically based research, but rather on his experience as a Navy flight surgeon.

An early study by Reinhardt (1970) summarizes some basic personality themes. This study focused on 105 naval aviators who were selected by their commanders as superior jet pilots. The aviators underwent two or more psychiatric interviews and a battery of psychological tests including the MMPI, Edwards Personal Preference Schedule, and the Maudsley Personality Inventory. Reinhardt (1970) noted the general characteristics of self-confidence, desire for challenge and success, little introspection, and a tendency towards interpersonal and emotional distance. There was also a significant number of firstborn sons with unusually close relationships with their fathers.

Studies done by Siem, Carretta, and Mercatante (1988) and Carretta and Siem (1988) used a wide variety of instruments to target expected attributes with 1,992 UPT students and graduates. The attributes included: Compulsiveness/Decisiveness; Risk-taking, Decision Making; Self-Assessment Ability, Self-Confidence; Survival Attitudes, Risk Taking; and Field Dependence/Independence. Despite the wide variety of measures and instruments, only the self-confidence measure appeared to contribute unique variance in predicting UPT success, beyond that explained by the AFOQT (Carretta and Siem, 1988).

Siem (1992), using the Automated Aircrew Personality Profiler with Air Force student pilots, found three characteristics related to training outcome but not improving the current selection model. The three factors were: hostility (negative relationship), self-confidence, and values flexibility (Siem, 1992). Similar limited results came from a naval flight training study using Aviation Qualification Test/Flight Aptitude Rating (AQT/FAR) and the Pilot Personality Questionnaire (PPQ) scores (Street et al., 1992b). In this study, the competitiveness measure was the most powerful predictor of overall training success. Likewise, a study with commercial aircrew found a pilot's performance rated by Check Airman could be predicted by "trait constellations of instrumentality and expressiveness as well as components of achievement motivation" (Helmreich, 1986, p. 276).

Gnan, Flynn, and King (1995) report that the German airline Lufthansa uses the Temperament Structure Scale (TSS) to screen applicants. The TSS measures 10 personality dimensions: work-related traits (Motivation, Rigidity, Mobility, and Vitality), social behavior traits (Extraversion, Dominance, and Aggressiveness), and stress resistance/emotionality factors (Emotional Stability, Spoiltness, and Empathy). Two unique and important dimensions are Vitality and Mobility. Vitality is designed to measure traits related to the physical demands of aviation such as long flights, unusual hours, and physical fitness. Mobility assesses the risking behavior of pilots in dangerous situations such as unusual situations or emergencies. Gnan, Flynn, and King (1995) describe a study of the TSS scales with 274 licensed airline pilots who were tested during their hiring process. Several dimensions of the TSS (Extraversion, Dominance, Emotional Instability, Aggressiveness) correlated with an airline job success criteria for three years after hiring.

Retzlaff and Gibertini's (1987) study of UPT students revealed three distinct clusters of traits measured by the Personality Research Form (PRF) and Millon Clinical Multiaxial Inventory (MCMI). In general, the pilots showed mainly histrionic and narcissistic features and almost no indications of severe personality disorders and clinical syndromes. The first PRF Cluster, 21% of the sample, was closest to a "right stuff" stereotype showing high aggressive, dominant, exhibitionistic, impulsive, and playful tendencies while being low on autonomy and self-direction. The second PRF Cluster, 58% of the sample, was one of high Achievement, Affiliation, Endurance, Social Desirability, and low Deference. Lastly, the third PRF Cluster which was 21% of the sample, exhibited low Affiliation, Change, Dominance, and Exhibition. Applying MCMI variables to the PRF clusters revealed similar differences. The first MCMI cluster had high Histrionic, Narcissistic, and Antisocial profiles. The second group had moderately Narcissistic and Histrionic, with high Compulsive, profiles. Finally, the third group was characterized by high Compulsive and low Histrionic profiles. The study concluded that the cluster differences could represent meaningful diversity in training performance or career selection and progression.

Picano (1991) completed a similar study using the Occupational Personality Questionnaire (OPQ) with experienced US Army pilots. Picano verified three similar distinct pilot personalities. However, he found no relationship between personality cluster and type of mission flown (i.e., attack, observation, utility). He concluded there was no one type of personality that was more or less suited to being a pilot or flying a specific mission. The only occupational relationship with the personality clusters was that a higher proportion of one cluster held instructor ratings which might be explained by the cluster's "competitive and achievement-oriented" drive (Picano, 1991, p. 520). Picano cautions that the personality clusters probably do not adequately represent the military pilot population because the sample was all volunteers. The OPQ is designed for use in work settings and is based on personality variables theorized to be important to work. The three main personality dimensions are relationship, thinking, and feeling. OPQ Cluster 1, 48% of the sample, could be labeled "methodical extroverts." This cluster is "the most affiliative and outgoing" and uses "a structured approach to problem solving which emphasizes planning, logical analysis, and attention to detail." OPQ Cluster 2, 36% of the sample, could be labeled "introverted worriers." This cluster showed traits of being "emotionally controlled, inhibited, apprehensive, and socially retiring" preferring "stability, security, and predictability." They seem to be "uncomfortable in social situations and pessimistic in outlook." OPQ Cluster 3, 16 % of the sample, could be labeled "competitive individualists" which correspond to the "right stuff" stereotype. These pilots were "highly independent, competitive, and decisive." They tended to be "least concerned with making a good impression and least emotionally sensitive and empathic" (Picano, 1991, p. 520).

Flynn, Sipes, Grosenbach, and Ellsworth (1994) tested F-16 fighter pilots anonymously with various psychological measures including the MMPI-2, Personal Characteristics Inventory (PCI), Computerized Diagnostic Interview Schedule (C-DIS), and MAB. The major goal of the study was to use a computerized battery to assemble normative psychological data from a representative sample of aviators. In this case, the pilots were volunteers from a squadron. Moreover, the anonymity of testing was considered to lessen the threat to pilots. From 23 pilots (64% of the squadron), the researchers found expected low scores on health complaints, depressive complaints, acknowledging stereotypical gender roles, and comfort in social situations. Likewise, their MMPI-2 scores were high for optimism and being active, outgoing, and energetic. The PCI scores for 20 pilots (42% of the squadron), measuring "crew coordination qualities," showed clusters of top eight, middle eight, and bottom four pilots. The top group was identified by goal seeking, achievement orientation, and interpersonal orientation. The bottom group showed higher verbal aggressiveness and low interpersonal orientations. In addition, unremarkable results came from the five pilots who took the C-DIS; the National Institute of Mental Health's (NIMH) epidemiological survey to screen for the prevalence of psychiatric disorders. Nevertheless, the researchers think that aviator C-DIS data could help prevention of mental health difficulties.

Longitudinal studies and more relevant psychological tests may offer predictive personality data otherwise unavailable. One such study, Neuropsychiatrically Enhanced Flight Screening (N-EFS), collects cognitive and personality baseline data from flight screening student volunteers (King and Flynn, 1995). The cognitive and personality data were gathered with several instruments: Multidimensional Aptitude Battery (MAB); CogScreen; NEO Personality Inventory-Revised (NEO-PI-R); and Personality Characteristics Inventory (PCI). Their goal is to track the students' progress beyond the traditional criterion of UPT success to operational success of becoming a mission-ready pilot. The NEO-PI-R, which is based on the five-factor model of personality, is purported to be more suited than previous personality measures to studying the normal range of personality functioning (Costa and McCrae, 1992). The NEO-PI-R's domains, the five factors of personality, include Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (see Appendix A for a delineation of the facets of each domain).

The five-factor model seems appropriate for aviators because of its global nature and ability to measure higher-functioning occupational attributes (Barrick & Mount, 1991; Helton & Street, 1992; Street & Dolgin, 1993; Tett, Jackson, & Rothstien 1991). Moreover, Costa and McCrae (1987, 1988, 1989) and Digman (1990) suggest that the five-factor model is the best general personality model describing the normal range of personality functioning.

CHAPTER IV

PILOT PERCEPTIONS OF EXPERT PERFORMANCE VARIABLES

Pilots can provide useful descriptions of desirable performance. Expert and experienced pilots are the operational subject matter experts. They exemplify operational performance and their opinions on what constitutes effective and safe performance can help set research performance standards. Several studies have found consensus among pilots on what aspects of performance are most important.

Siem and Murray (1994) asked 100 Air Force pilots from different aircraft to rank the importance of sixty traits of effective performance based on the Five-Factor Model. The Five-Factor traits included: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Culture. The personality trait of Conscientiousness was identified as the most important personality trait for combat performance across different aircraft and performance dimensions. The highest ratings associated with Conscientiousness were the trait factors of Discipline, Decisiveness, and Responsibility. The performance dimensions included: 1) Flying skills and knowledge; 2) Compliance with regulations and procedures; 3) Crew management and mutual support; 4) Leadership; 5) Situational awareness; and 6) Planning. The Conscientiousness dimension has also been found to be the most predictive personality dimension with five other occupations where correlations for subjective criteria were larger than for objective ratings (Barrick & Mount, 1991). Interestingly, the larger impact of subjective ratings suggest that the predictive value of some personality dimensions may be more dependent on subjective perceptions than actual objective criteria.

Murray, Siem, Duke, and Weeks (1995) used accounts of critical incidents during Desert Shield/Storm to yield six dimensions of *individual* pilot performance. These dimensions are: 1) Compliance with Regulations; 2) Knowledge, Skill, and Ability; 3) Crew management, utilization and mutual support); 4) Leadership; 5) Situational Awareness; and 6) Planning (Murray et al., 1995). These categories are translations of representative categories which describe continuums between examples of desirable and undesirable performance: 1) adherence to directives *versus* breaking the rules; 2) high knowledge and ability in flight *versus* procedural errors; 3) working with people *versus* poor communication; 4) takes charge *versus* quits doing the job; 5) ability to prioritize *versus* no situational awareness; and 6) preparing for all contingencies *versus* poor mission preparation (Murray et al., 1995).

Two critical incident studies illustrate different ways to approach interpersonal processes. The first is a study with C130 crews creating 7 dimensions of *aircraft commander interaction* with crewmembers: 1) facilitating teamwork; 2) responsibility/accountability; 3) motivating/disciplining crewmembers; 4) training/coaching crewmembers; 5) coordinating/directing crewmembers; 6) facilitating information flow; and 7) problem solving/decision-making (Hedge, Hanson, Siem, Bruskiewicz, Borman, & Logan, 1995). Each scale was given a label, a definition and three behavioral statements defining high, middle, and low effectiveness in the given dimension. The second study with Air Force tanker crews came up with five *team-level* performance dimensions: 1) Maintaining an atmosphere that Facilitates Teamwork; 2) Backing Each Other Up; 3) Coordination; 4) Group Problem Solving; and 5) Information Flow (Hanson, Hedge, Logan, Bruskiewicz, Borman, & Siem, 1996).

Flynn, Sipes, Grosenbach, and Ellsworth (1994) surveyed 29 volunteer F-16 fighter pilots out of a squadron flying unit of 36 pilots. The pilots chose the best formation lead and two wingmen pilots from their squadron and then described their important qualities from a list of categories. These categories were compiled from a NASA peer survey of astronauts and "top pilot" characteristics collected from past aces. The nine rating categories include: 1) General Knowledge; 2) Job Performance; 3) Stress Tolerance; 4) Leadership; 5) Group Cohesiveness; 6) Teamwork; 7) Personality; 8) Communication Skills and; 9) Aggressiveness (see Appendix B for category descriptions). The results showed that both lead and wingman pilots were seen as having high job performance, but lead pilots were also expected to have more leadership and stress tolerance qualities. In addition, these pilots demonstrated an ability to agree on top performers.

CHAPTER V

SAFETY STUDY VARIABLES

Safety may be a neglected, yet critical, aspect of performance. Many accidents are the result of poor performance. The goal of safety studies is to detect unsafe practices to improve performance and prevent accidents. Safety studies review accidents for contributing factors such as personalities and attitudes, the effects of stress and stress-coping, and common types of pilot errors.

Heinrich, Peterson and Roos's (1980) work on industrial safety management outlines some fundamental precepts of safety theory. All accidents are the result of unsafe acts and conditions. A review of industrial accidents showed that for a single major injury, there was an average of 29 minor injury and 300 no injury accidents. In addition, the sequence of events leading up to an accident proceed in a domino effect. Each event sets the stage for the next so an injury is the result of an accident which is the result of an unsafe act or condition. Therefore, if an element in this chain of events is removed, the accident process may be stopped. Safety management can focus on the more frequent and less severe events to prevent major accidents However, Alkov (1996) stresses that accident prevention is not as simple as focusing on single factors.

The two main categories of safety studies are observational and experimental (Li, 1994). Experimental studies involve manipulating factors in the event being studied where observation does not. Most of safety research is observational research based on case reports of accidents.

There are difficulties linking accidents to poor performance. Inadequate performance does not always result in an accident, and accidents are not necessarily caused simply by pilot performance. As Diehl (1991) points out, there are many hazards, of which only some lead to incidents and even fewer incidents lead to accidents. Due to this circumstance, research based on evidence from accidents provides a limited sample of actual poor performance.

In reality, accidents are caused by a wide range of factors and often a combination of these factors, not only pilot error (Baker, Li, Lamb, & Warner, 1995a; Baker, 1995b; Li, 1994a; Li & Baker, 1994b). Likewise, pilot factors may or may not play a critical role in an accident, although crash investigators are more likely to ascribe accidents to pilot factors rather than environmental factors (Baker et al., 1995a; Baker, 1995b). The likelihood of a certain pilot being in an accident depends on a variety of factors such as number and degree of hazards, training preparation and experience level, risk-taking propensity, psychomotor functioning, and an ability to recognize and respond to dangerous situations (Baker et al., 1995a).

Baker, Li, Lamb, and Warner's (1995a) review of twenty air taxi and commuter pilots who had repeat air crashes during 1983-1988 indicate that repeat crashes may be due more to frequent exposure to hazardous flying conditions than personal factors. There was little evidence suggesting that the major factors in repeat accidents were due more to pilot personality characteristics than single-accident pilots. In addition, the crashes of repeaters were twice as likely to occur in Alaska where there are more environmental hazards such as inclement weather conditions and unpaved runways.

The accident data may be limited by several factors, which makes the analysis of contributing causes difficult and potentially misleading. Post-accident interviews with pilots, peers, friends, and family are likely to be affected by many other issues. If the pilot is alive and will agree to screening, the pilot's testimony is often protected by both military and civilian aviation. Likewise, pilots are reluctant to disclose information that might put their aviation career in jeopardy. The worst case is when the pilot dies and there is no way to ask questions about the accident or prior history. Then the federal aviation and safety studies rely on limited information from the Cockpit Voice Recorder (CVR) and Flight Data Recorder.

Another source of safety-related information is NASA's Aviation Safety Reporting System (ASRS). The ASRS is a database of safety-related reports volunteered by pilots, controllers, and others for the purpose of identifying problems and solutions. The program encourages aviation personnel to contribute their accounts of dangerous events by guaranteeing confidentiality. These subjective reports offer rich details in spite of potential reporting biases (Degani & Weiner, 1990). ASRS has processed 338,000 aviation incident reports, issued more than 2,500 alert messages, accommodated 4,800 database search requests, and published 56 research studies in its twenty years of existence (NASA, 1996).

Personality

Safety studies try to find relationships between personality characteristics and the likelihood of being in an accident. Most safety studies start with hypothetical dangerous personality traits that they look for in accident histories. While there has been success in identifying dangerous personality traits, these traits are not confirmed by other studies.

Lester and Bombaci (1984) tested the five "hazardous thought patterns" used by the FAA and Transport Canada to illustrate faulty judgment. These irrational thought patterns are: 1) Anti-Authority: "Don't tell me!", 2) Impulsivity: "Do something- quickly!", 3) Invulnerability: "It won't happen to me!", 4) Macho: "I can do it.", and 5) Resignation: "What's the use?." Thirty five civilian pilots were tested for the presence of these thought patterns. Researches used scores on Cattell's Sixteen Personality Factor Questionnaire (16PF), measuring integration/self-concept, impulsivity, and superego strength, and on the Rotter Locus of Control (I-E) scale. The pilots were also given a forced-choice inventory ranking reasons for how they would respond to 10 flight scenarios. The five possible reasons reflected the different hazardous thought patterns.

Some significant relationships emerged in data analysis (Lester and Bombaci, 1984). Three predominant hazardous thought patterns were Invulnerability in 43% of the sample, Impulsivity in 20%, and Macho in 14%. Analysis of variance found the three hazardous thought patterns related to the 16PF integration/self-concept control scale and the Rotter Locus of Control scales. The results suggest that invulnerability, and to a lesser extent Impulsivity and Macho patterns, may be major mediators of irrational judgment and undesirable traits for aviation. The anti-authority and resignation patterns do not show statistical significance because they may be overlapped by the three predominant patterns or less common in the pilot population.

Platenius and Wilde (1989) created a 302-item questionnaire asking about life events, life styles, interests, and characteristics that might relate to being involved in an accident. This questionnaire was sent to 70,000 licensed Canadian pilots and approximately 12,000 were returned. The groups of pilot ratings were: 5,480 private, 1,969 commercial, and 1,084 airline transport certificates as well as 285 helicopter ratings. Using stepwise multiple discriminant analysis, the accident-markers were empirically re-categorized. These new categories were not psychologically definable but could retrospectively identify pilots who had accidents. Most of the markers kept by the discriminant analysis were psychological in nature and the rest were biographical and flying experience. Some of the accident markers for the different groups of pilots included: 1) preoccupations about business decisions and divorce or separation, 2) risk acceptance (flying regardless of what others say or being pressured to fly when they did not want to fly), 3) relatively asocial and sedentary hobbies, 4) perceived lack of success, and 5) automobile accidents and driving violations. Notably, this study did not find the characteristics of invulnerability, impulsivity, and macho to be clear accident-related factors.

Using the 16 PF, Mehrabian Achievement Scale, and a dynamic decision making task (under risk), Sanders and Hoffman (1975) found three personality factors that predict 86% of a sample of army aviators who were in military accidents. From a stepwise discriminant function, the three factors that significantly discriminated between pilots who had or had not been in error-involved accidents were: 1) Group dependent versus Self-sufficient, 2) Practical versus Imaginative, and 3) Forthright versus Shrewd. In general, pilots who had not been in an error-involved accident were measured to be more independent, creative, and direct.

Stress Coping

Stress-coping styles and comfort with risk are important personality dimensions affecting pilot performance. Researchers know much about the different forms of stressors and general reactions to stress. In general, too much stress can impede performance and increase potential for accidents. Therefore, many aviation studies examine the role stress plays in pilot performance as well as the coping abilities of pilots.

Latack and Havlovic (1992) define stress coping as "constantly changing cognitive and behavioral efforts to manage the internal and external demands of transactions that tax or exceed a person's resources"

(p. 483). Coping can be conceptualized as a by-product of a person-environment transaction due to a person's appraisal of harm, threat, or challenge. The term stress coping refers to both effective and ineffective responses to stress. These coping responses are usually problem or emotion focused. Likewise, the methods of coping can vary between: 1) cognitive versus behavioral, 2) proactive/control versus escape/resignation, and 3) social versus solitary. (Latack & Havlovic, 1992)

Stresses can come in many forms from present or past situations. Stokes and Kite (1994) describe three types of stress that affect aircrew performance. These types of stress are acute reactive stress, environmental stress, and life stress. The first type, acute reactive stress refers to a short term effect associated with operational tasks and situations such as workload and time pressures. Second is environmental stress which is the ambient physical conditions including noise, temperature and vibrations. Lastly, life stress is the accumulation of significant events in a person's life such as financial pressures and relationship changes.

The impact of stress is determined by the total stress and the individual's stress coping abilities (Alkov, Gaynor, & Borowsky, 1985). Raymond and Royce (1995) show that the total amount of stress can be viewed on a continuum of too little and too much. Stress or emotional tone is important for normal functioning. Too little stress results in sleepiness, decreased attention, and slower reactions. Too much stress can interfere with a pilot's ability to focus on and respond to situations, often resulting in missing or responding prematurely and unnecessarily to stimuli (Raymond and Royce, 1995). The amount of stress that can be tolerated often depends on the pilot's stress-coping ability.

With regard to stress coping, there are specific and enduring dimensions. Stokes and Kite (1994) distinguish between personality states and traits where states are acute reactions to specific situations and traits are more chronic and extensive characteristics of a personality. In addition, cognitive appraisals are dependent on a person's specific perceptions of a situation as well as general belief systems. These positive or negative expectations of the environment tend to operate on an unconscious level and can be especially influential in unfamiliar or ambiguous situations. Overall, there are different forms and degrees of stressors, different abilities to cope with stress, and different ways stress is perceived dependent on the individual and situation.

Selye (1974) identified three stages of stress: 1) alarm, 2) adaptation, and 3) exhaustion. In the alarm stage, the body prepares to fight by increased heart rate and adrenaline. The second stage of adaptation is where the person tries to resolve the stressor and return to normal body functioning. If the stressor is not dealt with satisfactorily, the body remains in a heightened state and approaches the final stage of exhaustion.

Much is known about the effects of stress in general. People under significant stress tend to regress back to primary modes of behavior (Stokes and Kite, 1994). The stress-related regression or dominant response has several implications for pilot performance research. Under normal circumstances, all pilots perform well but when things start going wrong, some pilots perform well and others' performance deteriorates. The quality of pilot decisions vary under different levels of perceived and actual workload and danger. In addition, responses to stress often lead to inappropriate prioritization of attention and actions. Perceptual and cognitive attention may be given to the most psychologically central or salient matter instead of important peripheral information. This narrowing of attention is called cognitive and perceptual "tunneling" (Stokes and Kite, 1990, p. 68)

Stokes and Kite (1994) list other ways stress can impact performance. Working memory or short term memory (STM) has been shown to be vulnerable to stress where long term memory (LTM) is more resistant to deterioration. Cognitive stressors reduce both the STM storage capacity and information processing strategy functions. On the other hand, studies with different stress applications showed that while problem-solving relying on STM was sensitive to stress, problem-solving requiring LTM declarative knowledge was unaffected by the stress (Wickens in Stokes and Kite, 1994). However, LTM is not completely unaffected by stress. Stress may have a negative affect on the encoding of information into LTM (learning) while retrieval of information from LTM (remembering) remains stable.

Besco (in Stokes and Kite, 1994) describes expected stress-resistant qualities of superior pilots; maintaining calm when problems arise, handling problems effectively, and being a stabilizing influence. The various descriptions of these qualities include: detecting mistakes immediately; handling errors gracefully; sharing error assessments with crew; expecting errors to occur and knowing that they can handle these errors; not letting errors interfere with their performance; the ability to resist personal or organizational pressures to test marginal conditions; provide stabilization when the system is unraveling or in conflict; and quickly adapting to changes in tasks or environment. These qualities represent an ideal of stress-resistance. This kind of stress resistance may be exemplified by University of Illinois simulation research finding that superior pilots appear not to let stress hurry their decision making. (Stokes and Kite, 1994).

Several studies suggest that pilots tend to cope with stress in predictable ways. Pilots are often dominant, action-oriented, not introspective, and have high needs for mastery and control (Alkov, 1996; Picano, 1990; Retzlaff & Gibertini, 1988). Pilots are perceived as coping well with stress and, because their job involves so much stress, actually seeking stress (Alkov et al., 1985). Picano (1990) confirms the significant stress-coping styles of pilots in comparison with non-rated military personnel. Using the Coping Orientation to Problems Experienced (COPE) with army pilots, Picano (1990) found that the pilots prefer "problem-focused strategies oriented towards direct action to master stressful situations" and "tended to de-emphasize emotion-focused forms of coping with stress" (p. 359).

Based on a review of mishaps, Alkov, Gaynor, and Borowsky (1985) conclude that pilot error can be a symptom of inadequate stress coping. This study hypothesized that typical young male naval aviators are aggressive and non-introspective. When these aviators experience stress, they "act out" internal tensions, which is evident in interpersonal relationships. Alkov asked naval flight surgeons, who served as members of mishap boards, to gather information on mishap pilots through interviews with squadron personnel and family members. Aviators who played a significant role in a mishap were compared with those who did not. The aviators who contributed to a mishap were more likely to "act out" (indicated by problems with interpersonal relationships, troubles with superiors and peers) and exhibit certain factors that would make them more vulnerable to stress. These factors include immaturity, no sense of their own limitations, and an inability to assess potentially troublesome situations (Alkov et al., 1985).

Similarly, Raymond and Royce (1995) offer ways to identify aviators at risk, who were previously labeled failing aviators (Voge, 1989). An aviator at risk may be over-stressed by extreme situational factors which, in turn, degrade performance and increase the risk for a mishap (Raymond and Royce, 1995). Alkov, Gaynor, and Borowsky (1985) emphasize the likely short-term and situational nature of this kind of pilot error due to inadequate stress coping. According to Raymond and Royce (1995), the warning signs of possible risk are: 1) Denial, defensiveness, over-sensitivity to criticism, 2) Argumentative, arrogance and/or hostility, 3) Interpersonal problems with bosses, peers, and spouses, 4) Financial problems, 5) Behavioral excess (e.g. eating and drinking), 6) Withdrawing socially, 7) Fatigue, 8) Deteriorating or poor flying performance, 9) Increased risk taking, and 10) Personality changes.

Researchers have also tried to develop psychophysiological stress-reaction profiles. Sive and Hattingh (1991) studied 17 Boeing 737 pilots' reactions to stressors in a flight simulator. The stressor used in this study was a birdstrike on takeoff after the plane was committed to continuing the takeoff. Psychological and physiological stress measurements were taken at different stages of the simulated flight. The psychological measure, conscious perceptions of anxiety, were measured with the State-Trait Anxiety Inventory. Plasma cortisol, catecholamines, lactate, total protein, osmolality, total lipid, glucose, and haematocrit were selected as physiological measurements due to evidence that they reflect physiological reactions in certain animal species. The various chemicals displayed different reactions to the stressor which suggest that they would have future use in predicting stress. Some of the complications using physiological reactions were differences in body chemical compositions due to age differences and unpredictable changes in chemicals due to systemic factors other than stress.

Sive and Hattingh's study suggests several important points. First, it determines that there is potential to integrate physiological and psychological measures to get a fuller picture of how a person reacts to stress. Secondly, it proposes that the degree of stress awareness is an important consideration when interpreting the State Anxiety profiles. For example, the captains may have repressed anxiety before the emergency. Only when the captain knows what the emergency is and has control of the situation, will he/she be able to allow more awareness of anxiety and, ironically, score higher on an anxiety profile. In addition, there may be different types of stress depending on the person's perspective such as anticipating before or reacting after an event occurs.

Pilot Errors

Pilot errors are examples of poor performance. The NTSB (1994) defines pilot error as "a discrete instance in which a crew member (1) did something that should not be done, (2) did something inadequately, or (3) did not do something that should have been done" (p. 9). The different kind of errors are useful in discriminating how performance breaks down and what danger results. Most of these error trends are culled from safety studies. The trends of pilot errors tend to be very consistent across time and airframes. The detail and scope of error analysis is improving as the quantity and quality of available accident data increases.

An early study by Cooper (cited in Gregorich et al., 1990) studied 60 accidents that occurred between 1968 and 1976 where crew coordination played a major role. He found that deficiencies arose in common areas of: 1) inappropriate amounts of attention given to minor problems, 2) leadership, 3) delegating tasks and assigning responsibilities, 4) setting priorities, 5) monitoring crew and aircraft systems, 6) using available data, and 7) communicating intentions and plans.

In a more recent and comprehensive study, the NTSB investigated 14 factors in their review of flightcrew-involved, major accidents between 1978 and 1990. These factors include: 1) type of operation; 2) phase of flight; 3) flight delay status; 4) equipment type; 5) crew member position and function; 6) workload of the crew member in relation to the quality of information available to the crew member when an error occurred; 7) fatigue; 8) fitness; 9) stress; 10) past performance evaluations; 11) mutual familiarity of the crewmembers; 12) training; 13) experience; and 14) air carrier organizational structure and function (NTSB, 1994). The study explains that the definition of error is restricted by what kind of information can be obtained reliably after an accident. A limitation of accident studies, in general, is that errors in perception, comprehension, attention, knowledge, memory, or reasoning, which may have led to an error of action or inaction, are difficult to determine after an accident.

The NTSB safety study also classifies three types of action/inaction errors. These classifications are: 1) Aircraft handling (failing to control the airplane in desired parameters); 2) Communication (Incorrect readback, hearback; failing to provide accurate information; providing incorrect information); 3) Navigational (Selecting wrong frequency for required radio navigation station; selecting the wrong radial or heading; misreading charts); 4) Procedural (Failing to make required callouts, making inaccurate callouts; not conducting or completing required checklists or briefs; not following prescribed checklist procedures; failing to consult charts or obtain critical information); 5) Resource management (Failing to assign task responsibilities or distribute tasks among crewmembers; failing to prioritize task accomplishment; overloading crewmembers; failing to transfer/assume control of the aircraft); 6) Situational awareness (Controlling aircraft to wrong parameters); 7) Systems operation (Mishandling engines or hydraulic, brake, and fuel systems; misreading and mis-setting instruments; failing to use ice protection; disabling warning systems); 8) Tactical decision (Improper decision making; failing to change course of action in response to signal to do so; failing to heed warnings or alerts that suggest a change in course of action); and 9) Monitoring/challenging (Failing to monitor and/or challenge faulty action or inaction by another crewmember. The first eight errors are primary errors and the last error is a secondary error because it is dependent on another crewmember making a primary error.

Diehl (1991) specifies a different conceptualization of three possible types of errors. These types of errors are: 1) Procedural "slips," 2) Perceptual motor "bungles," and 3) Decisional "mistakes." Procedural errors deal with mismanagement of the aircraft systems and configurations. Perceptual-motor errors are improper inputs to power and control surfaces. Decision errors are poor judgments planning the flight and evaluating the conditions. Using these error types, Diehl examined airline and scheduled airtaxi accidents involving aircrew error and incurring fatalities and/or destroyed aircraft between 1987 and 1989. Aircrew error was present in 24 of these 28 major accidents. There were 16 procedural, 21 perceptual-motor, and 48 decisional errors. Diehl also did the same type of analysis on Air Force mishaps which resulted in destroyed aircraft, over one-million dollars damage, and/or fatalities. Aircrew error was found in 113 of 169 (67%) mishaps. In the Air Force accidents, there were 32 procedural, 110 perceptual-motor, and 157 decisional errors. According to the relative numbers, decisional errors account for a majority of the total aircrew errors.

Baker, Lamb, Li, and Dodd (1993) studied the human factors in commuter crashes during 1983-1988. The majority of the accidents involved inadequate pilot performance. In order of frequency out of 118 events, the pilot performance errors involved: Emergency handling (25), IFR procedures (11), Fuel management (9), See and avoid procedures (8), Preflight procedures (7), Judging weather and terrain (7), Hazardous runway conditions (6), Landing gear configuration (6), Handling airport wind and turbulence (4), Judging short and/or narrow runways (2), and Weight and balance (2). In addition, there were 30 cases where there was no obvious pilot factor.

In her study of commuter crashes, Baker (1993) draws several conclusions. First, deficiencies in routine and emergency procedures show a need for adherence to existing procedures and improvement of those procedures. Second, similar breaches in checklist discipline and poor analysis are also evident in the improper handling of emergencies, where problem solving is hurried and leads to erroneous conclusions and actions. Lastly, there is a high number of errors due to improper Instrument Meteorological Conditions (IMC) knowledge and procedures. These problems range from insufficient understanding or compensation of the affects of various weather conditions to incorrect instrument set ups and imprecise altitude and airspeed control.

CHAPTER VI

CREW RESOURCE MANAGEMENT VARIABLES

Crew Resource Management (CRM) training and research uses social psychology and management theory to understand and improve cockpit interactions (Diehl, 1991b). The main thrust of CRM is training crews to be more effective and safer. Usually, CRM programs use simulators to create flight-like environments and tasks where crews can practice how they handle various situations. After simulator flights, crews debrief the mission and review video tapes of their performance. The airlines commonly refer to this type of training as Line Oriented Flight Training (LOFT) and the military generally call their programs Mission Oriented Simulation Training (MOST) (Helmreich, 1986) or Aircrew Coordination Training (ACT) (Alkov, 1994).

CRM brings together a wide range of important performance factors. For instance, the Hughes Training Crew Resource Management Workbook covers policy and regulations, command authority (leadership), aircrew communications, workload performance, available resources, situational awareness, decision making, and operating strategy. The purpose, importance, and typical types of mismanagement of each factor are discussed as well as tools for safer and more efficient management.

Research utilizes CRM models and outcomes to measure different aspects of performance. CRM is a costly program that is implemented to reduce the number of crew-related accidents. The primary goal of research is to validate if CRM training actually decreases errors and accidents. These studies use simulator experiments and accident data to determine whether pilots who have received CRM training make fewer errors than other pilots. Simulators offer control and standardization of testing environments and tasks. In comparison, changes in accident rates provide operational evidence of improved performance. Both types of research indicate that CRM is effective in reducing errors and accidents. Consequently, CRM is probably a valid model for studying pilot performance.

Pilot performance studies look at different personalities, management styles, and communication styles using CRM profiles and standards. The key issues are self and interpersonal awareness and interpersonal abilities such as communication and leadership (Skogstad, Dyregrov, & Hellesoy, 1995). Therefore, close attention is put on group patterns of leaders and team members. These factors are studied in relation to how the basic processes are handled. These fundamental processes are similar to Diehl's (1991b) five cockpit management tools, which are: 1) Attention management, 2) Crew management, 3) Stress management, 4) Attitude management, and 5) Risk management.

Leadership is a critical part of successful aircrew performance. Flight leadership manages crew resources to achieve mission goals safely and effectively. Resources are managed by providing structure and direction. The ideal amount of structure is enough structure where there is certainty about what is expected from each crew member but not too much where each crew member is limited in taking initiative and responsibility. Likewise, goals let the crew members know where they are going. Providing clear structure and goals can give crew members the security to concentrate on their job instead of questioning what their job is or why they do the job. In many ways, a good leader is like an effective parent who provides appropriate structure and freedom. Effective parents have been shown to be "authoritarian" as opposed to the extremes of being "permissive" or "autocratic" (Baumrind, 1970).

Chidester, Kanki, Foushee, Dickinson, and Bowles (1990) studied how different task-oriented and relationship-oriented leadership profiles affected simulator performance. Using a variety of instruments with airline captains, Chidester and colleagues found three basic pilot personality clusters which were evaluated by relative performance on simulated missions. The three clusters were: 1) Positive Instrumental-Expressive (IE+) Profile, 2) Negative Expressive (EC-) Profile, 3) and Negative Instrumental (I-) Profile. Instrumentality represents goal orientation, independence, and "achievement striving" (how a person cares for the job). In comparison, expressivity represents interpersonal warmth and sensitivity (how a person cares for others). The IE+ captains displayed high achievement motivation and interpersonal skill. The EC- captains tended to have below average achievement orientation and a negative expressive style such as complaining. Lastly, I- captains were more likely to be competitive, verbally aggressive, impatient, and irritable. These characteristics were measured with the Expanded Personal Attributes Questionnaire (EPAQ), the Work and Family Orientation Questionnaire (WFOQ), and the Achievement Striving and Impatience/Irritability scales (A/S, I/I). Performance was evaluated on performance scales by an expert evaluator who sat in the simulator and independent raters using video recordings of the flight. This study found that the crews led by a IE+ captain were consistently effective and made fewer errors. Crews led by EC- captains were consistently less effective and made more errors. However, crews led by I- captains were less effective on the first day but equal to the best on subsequent days.

Hedge, Hanson, Borman, & Bruskiewicz, 1994) used a Situational Test of Aircrew Response Styles (STARS) to measure CRM skills. This computer test uses crew resource management incidents and response options to evaluate and predict performance. Using job-relevant situations, the situational judgment test asks which alternatives would be most and least effective in a given situation. The assumption is the differences in answers should reflect the subtle differences between effective and ineffective judgment.

Hedge et al. (1994) claim paper and pencil situational judgment tests, in general, offer several advantages. These tests are inexpensive because they can be administered to large groups simultaneously. Also, there is low to moderate correlations with standard general ability and academic achievement measures, the situational judgment test can provide incremental variance beyond these standard measures. There are little differences in scores between genders and different racial groups. Since the test items are created from actual job scenarios, they are likely to appear more relevant and attractive to test takers. These tests may measure, what is sometimes called "street smarts," the practical abilities necessary to deal with real-world situations.

Ruffel Smith (1979) showed that effective delegation can be a critical skill. Since humans tend to narrow perceptual attention under stress, the decision-maker should avoid assuming too many tasks such as trying to fly, make decisions, and direct the crew. Captains who give the flying responsibility to the first officer benefit by making better decisions. In addition to creating more time and mental resources to make decisions, the other crew members feel a sense of responsibility. This scenario is in contrast to the situation where crew members feel they are not needed or wanted and become less involved.

The quality of communications can have a large impact on performance. Ironically, communication is often viewed as just a vehicle for messages instead of an essential means of group information processing. From an information processing perspective, communication represents the critical ways information is shared and activity is coordinated in the cockpit. Effective communication can improve situational awareness, decision making, judgment, workload management, and even stress management. The synergy of a group is much stronger than the sum of the individual members. On a crew aircraft, each individual has responsibility for communicating effectively because s/he fills a vital role and may see or remember something that would otherwise be missed. Each crew member is "another set of eyes and ears" as well as a mind with unique knowledge, skills, and perceptions. Individual differences can be turned into group strength through effective communications.

Effective communication entails a shared responsibility for sending and receiving messages in a constructive manner. The goal of communication is to have the listener understand a message and respond appropriately. Communication is a shared responsibility, requiring each member to be aware of, and appreciate, each other's thoughts, feelings, and behaviors (Alkov, 1994). A mission performance study (Foushee and Manos in Alkov, 1994) showed that effective aircrews use more frequent, direct, concise, and open communication by establishing an environment where people are comfortable and encouraged to share ideas and make suggestions and counterproposals.). Effective aircrews are also characterized by homogeneous speech patterns which enhances predictability of crew member behavior (Kanki, Lozito, & Foushee, 1989a; Kanki & Foushee, 1989b). Establishing a "regularity" to communications may establish an important underlying teamwork rhythm as well as helping crew members understand and rely on each other better. Effective communication also avoids unfamiliar, ambiguous, careless, or complex language that may confuse or give the wrong message to others. Foushee and Manos (in Alkov, 1994) also found that less effective crews had more disagreements and exhibited more "uncertainty, anger, frustration, and embarrassment"(p. 15).

Thinking in personal, instead of objective, terms can impair communication. A common example is when a person has a hidden agenda and "makes decisions and gives advice based on information or personal reasons unknown to the listener" (Alkov, 1994, p. 15). A person may also give incomplete information to avoid conflict or looking bad. Whether information is distorted or withheld, other people in the crew relationship are forced to make decisions based on inaccurate or incomplete information which leads to poor decisions. Extremes of poor communication impair reality testing and create confusion and distrust.

Communication has been studied using the sequence-variable concept where speech is divided into a two-part sequence of "initiating" and "responding" functions (Kanki et al., 1989a). In this study, initiating speech was further classified into: 1) Commands, 2) Questions, 3) Observations, and 4) Dysfluencies (e.g. ungrammatical or incomplete utterances, talking to oneself). Responses were interpreted as: 1) Any reply greater than a simple acknowledgment, 2) Acknowledgments, and 3) Zero response. The types and numbers of each communication sequence was analyzed with respect to crew position. In low error crews, the aircraft commanders gave more commands and the flight officers responded with more acknowledgments than expected. Overall, it appears that any communication convention or regimen improves overall performance.

Another critical aspect of CRM is following Standard Operating Procedures (SOPs). Deviation from established policies and regulations are involved in approximately 20 % of all mishaps according to Hughs Training Crew Resource Management Workbook. Alkov (1987) emphasizes that crew members can avoid mishaps by flying "by the book" all the time, especially during increased operational demands. Although increased operational demands are one of many temptations to deviate from SOP without necessity. These temptations vary from distorted thinking that the rules are silly and everyone does it to real or perceived pressures from others or even from the pilot's desire to show he/she can do the mission.

Checklist usage is a tangible way to study how a crew adheres to procedures under different circumstances. Degani and Wiener (1990) outline the purposes and vulnerabilities of checklist usage through observations of line and simulator flights, interviews with line pilots and officials from federal aviation agencies, information from aircraft and avionics manufacturing companies, and incident/accident reports from the ASRS, NTSB, and ICAO. Checklists back up the pilots fallible memory as well as generate and coordinate the complex and time-critical tasks at different stages of flight (Degani & Wiener, 1990). Checklists also standardize crew communication and activities for maximum efficiency during high workload and stressful conditions. There are three steps to checklist usage including the initiation, the calls and responses, and the completion. Each of these phases is important to crew coordination and is susceptible to different distractions. These distractions may result in checklist errors or omissions.

Linde and Goguen (1991) analyzed required speech acts of performing checklists and responding to radio calls during simulated flight. With regard to two types of checklist interruptions (radio calls from outside the aircraft or crew conversation from inside the aircraft), the study examined two possible responses (interrupting the checklist or overlapping the other interruptions with the checklist) with regard to safety performance. Safer crews demonstrated better checklist continuity by a higher ratio of checklist speech acts to total speech acts during checklist performance. In addition, they were able to minimize the length of checklist interruptions.

CHAPTER VII

HUMAN FACTORS VARIABLES

Human factors variables utilize a process or systems orientation. This systems orientation focuses on the interactions of different elements in a given system. These variables are human factors constructs of information processing with complex tasks and environments. These information processing demands are often synonymous with job complexity (Hunter, Schmidt, & Judiesch, 1990). The constructs are also called mental models of information processing and represent very complex and non-linear processes (Cowan, 1988) which can be generalized but not fully explained. Researchers attempt to measure mental models with empirical models, analytical models, and verbal/written reports (verbal protocol or "thinking out loud") (Rouse & Morris, 1986)

Rouse and Morris (1986) suggest that the "black box" of mental processing models will never be fully understood and depicted but can be generalized. The common themes of manual/supervisory control and cognitive science mental models are describing, explaining, and predicting. Mental models are functionally defined as "mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning, and prediction of future system states" (p. 351) where the purpose is why a system exists, the function is how a system operates, the state is what a system is doing, and the form is what a system looks like. In this conceptualization, a mental model describes purposes and forms, explains functions and states, and predicts states.

Information processing and complex jobs share similar characteristics of many different tasks and resources, of variable reliability and importance. These resources are sources of information and assistance in the aviation environment varying from the pilot's perceptions and cognitive abilities as well as other people such as crew members and ground personnel. These variables are more in the theoretical, exploratory stage and are being developed as measures. These variables could be integrated to measure a personal and interpersonal information-processing or resource management profile.

A general information processing (IP) model is proposed by Hogan and Broach (1990) describing the stimulus-process-result (S-P-R) performance domain. Some of the components are the task and environmental inputs, short-term (working) and long-term memory, a central processor, and selected response outputs. Working memory temporarily stores sensory information, whereas long term memory stores domain and procedure knowledge. Domain knowledge is a structure of specific factors and events in a network of propositions. Procedure knowledge stores the production rules for using domain knowledge. The central processor works as an "inference engine" (p. 3) using long-term memory resources to encode and analyze incoming stimulus.

All of these processes involve perceiving, organizing, and utilizing information depending on the nature of the context. In general, the immediate demands are met with a quicker response and the complex demands are met with more deliberate cognitive approaches. The experts tend to be characterized by more comprehensive considerations and effective use of personal and interpersonal resources.

There are other similarities between the various information processing models. The different information-processing variables have different normative and descriptive models (O'Hare, 1992). The normative models tend to be abstract, theoretical descriptions which suggest the ideal way a task and environment would be approached. These models offer useful ideas for evaluating how errors can be made. On the other hand, the descriptive models describe what is seen naturalistically, with real tasks in the actual environment, which is often very different from the theoretical models. Empirical evidence shows that the descriptive models describe how pilots actually process information.

Each of these information-processing approaches can have different conceptualizations based on whether a demand is simple and immediate or more complex and time intensive. For example, situational awareness has been conceptualized as both near-threshold reactions (Secrist and Hartman, 1993) or more complex analyses (Endsley 1994, 1995a, 1995b). In the same way, judgment and decision-making are viewed on a continuum between decision time and cognitive complexity (Jensen, 1982).

Even the reliance and use of short and long term memory are similar for these processes. Situational Awareness, decision-making, and judgment are shown to be heavily dependent on large and accurate memory abilities (Adams, 1993; Endsley, 1995a). Likewise, the different memory strategies of short term chunking and long term schemas are reflected in each processing dimension (Adams, 1993; Endsley, 1995a).

The human factors models include situational awareness, decision making, judgment, workload, and task and resource management. These models overlap functionally yet each also stresses an important aspect of performance. Each model is discussed with respect to its definition, supporting tenets, and relation to performance.

Situational Awareness

Situational awareness (SA) is defined in several ways. One definition holds that SA is "adaptive, externally directed consciousness" (Smith & Hancock, 1995, p. 137) or a pilot's understanding of the state of the aircraft, its systems, and environments (Adams, Tenney, & Pew 1995). More technically, SA describes the dynamic interface of the pilot and environment where the pilot adapts by managing knowledge and behavior to achieve goals while accounting for the conditions and constraints imposed by the task environment (Smith & Hancock, 1995). SA empowers the agent with the available outside information and inside knowledge to respond to dynamic situations. Likewise, SA is an appropriate awareness of the things in the environment that are important, but not necessarily of other irrelevant environmental factors.

SA can be critical to the safe and effective operations of aircraft. Having SA better prepares a crew member to deal with expected and unexpected eventualities. On the other hand, the opposite of having SA is when a pilot is not aware of all available pertinent information and is more vulnerable to making mistakes. The potential mistakes are oversights, hasty inferences, forgetting tasks in queue, or any other way decisions may be based on incomplete knowledge or information. Crew members are vulnerable to these dangers under high workloads and temporal compression as well as during complacent reliance on automation. A large part of SA is managing interruptions and task-unrelated events where real-world events do not follow organizational principles other than occurrence or discovery (Adams et al., 1995).

A sequential description of SA is given in three levels that build on each other; basically perception, understanding, and prediction (Endsley, 1995a). The first and fundamental level is where relevant information is perceived. Level two is where the different elements of perceived data are assigned meaning in relation to operator goals. Lastly, level three is future events and system states can be predicted based on perception and understanding of relevant information. Endsley sums up SA as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p. 36).

In contrast to lengthy cognitive processes, Hartman and Secrist (1991; Secrist and Hartman, 1993) conceptualize a temporally limited and stimulus-bound SA as near-threshold information acquisition used by combat pilots. Based on the importance of SA in combat, these studies hypothesize six skills essential to SA, stressing the value of quick and accurate acquisition of performance-critical cues. Secrist and Hartman (1993) highlight the importance of this special form of SA and showed that target visual access time could be improved with training. Using computer monitor visuals and joystick controls, subjects detected, recognized, and identified masked and fleeting targets.

A potential measure of SA is the Situational Awareness Global Assessment Technique (SAGAT) (Endsley, 1995b). The SAGAT offers a comprehensive measure of the three levels and various elements of SA. The test is administered on a computer so it can be standardized while offering the flexibility to be tailored to specific systems and applied to other types of systems. Queries are given randomly to test as many SA dimensions as possible.

The results rely on the subjects approaching the test in a certain way. Subjects are directed to perform tasks as they normally would and consider the SAGAT inquiries secondary. However, subjects are encouraged to respond to all queries because: 1) information may appear unimportant but have at least secondary importance; 2) there may be information that makes a question very important; and 3) lower priority questions are included to avoid inadvertently providing artificial cues about the situation that will direct their attention when the simulation is resumed (Endsley, 1995b).

There are several other possibilities for measuring ability to maintain SA including: physiological techniques, performance measures, global measures, external task measures, imbedded task measures, subjective techniques (self-rating or observer-rating) and questionnaires (posttest, on-line, or freeze technique). Each of these possibilities has different advantages and limitations. The key reliability and validity considerations governing choice and use of the metric are: 1) measures the construct and not other processes; 2) discriminates in the form of sensitivity and diagnosticity; and 3) not affecting the construct by altering behavior and biasing data. (Endsley, 1995b)

Safety studies have identified causes and symptoms of loss of situational awareness (Alkov, 1994). These causes and symptoms can occur for individual pilots or crews. The causes identified by Alkov include: 1) Distraction, fixation, or preoccupation and failure to detect important cues; 2) Ambiguous, confusing, or unclear information; 3) Work under- or overload; and 4) Poor communication. Some symptoms are: 1) Complacency and a contempt for hazards; 2) Euphoria; 3) Ignoring uncomfortable feelings about a situation (One of a pilot's most reliable cues is a "gut feeling"); 4) No one flying the aircraft or looking outside; 5) Failure to meet targets (e.g., for airspeed, rate of climb, power settings, checkpoint times);and 6) Departures from standard procedures and violations of regulations that lead pilots to exceed safe operating limits. (Alkov, 1996, p. 12)

Frederico's (1995) recognition experiments reveal differences in how expert and novices differ in situational awareness. The research found only one difference between experts and novices: experts depend more on context than novices do. Experts, but not novices, rely on being aware of contextual elements when assessing appropriate schemas from past experiences. Interestingly, experts and novices score similarly on all other measures including the number of schemata, the scenarios per schema formed, the access avenues ascribed for these schema, the depth of scenario analysis, and the importance placed on conceptual and perceptual aspects of a problem.

Decision Making

Decision-making is "the ability of a pilot to respond to cues from the environment, evaluate the situation, come to conclusions, and act on those conclusions" (Adams, 1993, p. 1). Decision-making integrates sub-categories like prioritization, task management, and problem solving. Likewise, decision-making has been studied from similar perspectives including expert/novice recognition, task shedding and scheduling, workload management and decision biases, strategies in a complex environment, components of expert decision making, and models of decision-making. All of these studies aid in understanding and evaluating how decisions are made and how various factors improve or impede decision-making.

Alkov (1996) posits that decision making relies on effective risk management and judgment. Risk management is a continual effort of assessing and weighing the amount of risk associated with situations. If risk assessment is inaccurate, then decision-making and judgment will be affected. Decision making is the specific problem-solving abilities of collecting relevant information, defining the problem, and solving the problem in a timely and logical manner, whereas, judgment is the general ability to make safe decisions (Alkov, 1996).

Decision making can be viewed on two levels; static capacities and dynamic processes. Some of the static capacities are working memory, logical reasoning skills, spatial ability and cue sampling skill. The processes represent how these capacities are used to make decisions.

The two basic decision-making processes are analytical and pattern recognition. The analytical models are the normative and laboratory versions of an optimal process where exhaustive analytical calculations are made. Analytical models are "bottom-up" approaches that consider all alternatives searching for the best solution. In contrast, pattern recognition models are more descriptive of how people actually make decisions in a natural setting which can be very different from the normative, analytical model. Pattern recognition is a "top-down" approach that settles on the first satisfactory fit between long term memory (LTM) schema and perceptual cues. This "top down" pattern recognition is thought to be how experts make decisions (Stokes & Kite, 1994).

Wickens, Stokes, Barnett, and Davis's (1988) normative model theorizes that decision-making is a utilitarian choice between the expected outcomes of all viable courses of action. The expected outcomes are the sum of the probabilities (expected frequency of occurrence) and utilities (positive or negative consequence) of all the possible consequences for each course of action. The optimal decision-process picks the course of action with the most favorable expected outcome and the lowest expected risk.

In comparison, O'Hare's (1992) descriptive model describes the actual decision-making process as different from the theoretical ideal. Experts sort and solve sequentially from most plausible explanation in their LTM memory schemas. If the most plausible explanation has inconsistencies, they proceed to the next most plausible using successive refinement.

A similar process of pattern recognition is observed in studies on expert thinking (Adams, 1993; Frederico, 1995). The concept of pattern recognition replicates Adam's (1993) description of cycling through schemas and Frederico's (1995) account of "schema-driven" recognition and decision-making. Meanwhile, the novices rely on a more mentally cumbersome analytical models until schemas are constructed (Frederico, 1995). However, Wickens, Stokes, Barnett, and Davis (1988) add that hypothesis generation and testing depends heavily on the mental "workbench" of working memory.

Expert schema-driven thinking compares experienced patterns and processes from memory with the experience of new situations. A critical part of the schema-driven thinking process is classifying same, similar, or different aspects of a situation by looking at context elements. This is probably why experts have been shown to be more context-dependent than novices (Frederico, 1995). Therefore, on simpler, more straightforward tasks, experts may perform equal to or even more poorly than novices who are using the most immediate hypotheses and solutions. But on complicated tasks, the expert process of taking more context cues into account and evaluating their importance should pay off with consistently better decisions. An analogy can be made to the spiral omnibus test construction where performance is differentiated by later questions containing attractive distracters (Anastasi, 1988). The novice may more readily accept solutions that appear to work.

The lack of performance variance caused by information-processing variables suggests that there are probably expert decision processes that are not being measured (Wickens et al., 1988; Frederico, 1995). Expert decision processes are harder to measure because experts probably create "specialized production rules" providing more efficient solutions based on previous experience (Frederico, 1995). Perhaps this is why certain information-processing variables can predict novice but not expert performance (Wickens et al., 1988). Also, expert processes probably involve meta-processes to monitor decisions for erroneous assumptions, expectations, and perceptions as well as evaluate the accuracy and importance of inputs from the environment and memory.

Adams (1993) uses the acronym SOARing (Sensing, Organizing, Analyzing, Responding) to summarize the Preparation and Execution phases of how expert pilots think. First, the Preparation phase utilizes Sensing (not only "sensing" information but also "attending to" relevant information where expert pilots pay attention to more details) and Organizing (filtering, prioritizing, and structuring sensed information using short-term "chunking" memory and long-term "schema" memory resources). Second, the Execution phase relies on Analyzing (information processing and evaluation where expert pilots use superior memory organizational capabilities which facilitates recognition and recall) and Responding (the most critical step where an action is taken to alter or control the situation and then monitor the effectiveness of the action). Three basic limits on decision-making are attention span, short term memory and long term memory (Adams, 1993).

Expert decision-making is characterized by superior memory, goal orientation, fast access, opportunistic planning, adaptability, self-monitoring, and perceptual superiority (Adams, 1993). Similarly, problem-solving research shows experts having superior knowledge structure; pattern perception; performance speed and accuracy; memory capacity; problem perception, representation, and analysis; self-monitoring skill; schema quantity; and context-dependency (Frederico, 1995). Adams (1993) concludes that memory is the fundamental process that accounts for differences in expert and novice pilot thinking abilities. In addition to large memory capacities, Adams (1993) explains "the autonomous information processing of many of their skills frees-up greater storage" (p. *x*). A goal orientation organizes concepts by procedures for their application and conditions (contexts) where the procedures apply. Fast access is evident in faster skill-based tasks which frees up working memory for processing other aspects of a problem as well as avoiding an extensive search of memory. Expert pilots use "opportunistic planning" to adapt conventional production rules (the flying and problem solving procedures for normal and abnormal situations) to a given situation while simultaneously evaluating multiple possible interpretations of a situation. Adaptability is a step above the routine expert, where there is a creative capability to respond to situations with limited or ambiguous information. Self-monitoring means being able to estimate problem difficulty and divide time among tasks accordingly. Perceptual superiority means a rapid ability to recognize and recall large meaningful patterns from a large knowledge base (Adams, 1993). Through experience, the pilot develops associative problem solving capabilities and thereby a capacity for more dynamic thinking.

Effective decision making in a time-limited, complex environment must assess "what and when" to determine how important a task is and when it should be addressed relative to other tasks (Raby and Wickens, 1994). Heuristics and "what and when" biases can be a decisional crutch as well as a handicap. Decision making in complex systems uses biases and heuristics to simplify decision making but often involves costs of unwarranted assumptions and missing important details causing poorer performance (Raby and Wickens, 1994).

Research has shown that people do not always make decisions optimally because of certain biases and heuristics (Wickens et al., 1988). Overall, people are deficient in generating and evaluating problem definitions and solutions while being overconfident in these abilities. People also tend to misassign the relative importance of perceived information based on considering all information as equally valuable or the most prominent information as the most important and likely.

First, people do not generate all the possible hypotheses (potential conceptualizations of the problem) and courses of actions (potential solutions) for a given situation, nor are people skilled at "assessing the probability of different outcomes and their resulting risks" (Wickens et al., 1988, p. 12). Second, people tend to be *overconfident* in all aspects of decision making. People are overconfident that they can generate a comprehensive list of hypotheses. People also tend to be overconfident about abilities (speed and accuracy) and schedule too many tasks or delay tasks until it is too late (Raby & Wickens, 1994). People even tend to overestimate the probability of their future expectations which can be viewed as overconfidence, the "can do" and "it won't happen to me" pilot biases, and an inherent dislike of uncertainty.

Third, people assume that their hypotheses and risk assessments are the most probable (accurate) because of the *availability heuristic* which illogically focuses on salience and accessibility. For example, people consider a hypothesis more probable because it is more accessible (most recently experienced) or salient (sticks out) from memory. The availability heuristic can also inflate expectancies, so people expect frequently occurring or popular positive and negative events. Similarly, there is a *salience bias* or a general tendency for people to focus on what is most salient in the environment. People tend to pay attention to cues that are loud, bright, recent, centrally visible and easy to interpret (Wickens et al., 1988).

Fourth, once a hypothesis is formed, a *confirmation bias* can influence cue seeking behavior to look only for evidence supporting the hypothesis the person already believes to be true. Fifth, the evaluation of cue reliability and task importance may be obscured by the *"as if" heuristic*, assuming all cues and tasks are equally relevant. The "as if" heuristic gives all information sources or tasks equal importance instead of prioritizing (Raby & Wickens, 1994). Sixth, there is a *framing bias* in which a choice between a guaranteed outcome and a risky option will depend largely on the framing of the options. According to a study done by Tversky and Kahneman (1981), if the choice is framed as between losses, people are likely to choose the risky loss rather than a certain loss. On the other hand, when choosing between gains, people are more likely to choose the guaranteed positive outcome.

Pilots have internal and external mechanisms to avoid the dangers of heuristics. Some of these are heuristics themselves. Two common pilot task-prioritizing heuristics are "aviate, navigate, and communicate" and "maintain aircraft control, analyze the situation, and take proper action." Likewise, pilots are highly methodical and create acronyms to use at mission milestones to remind themselves of all appropriate tasks. Some examples of these acronyms are the WANTS (Weather, Alternate, NOTAMS, TOLD, SID) check and the six T's crossing fixes (Time, Turn, Throttles, Twist, Track, Talk). Pilots also monitor their own decision-making by playing devil's advocate and questioning their assumptions. These questions usually start with "what if?" and "why?".

Standard operating procedures (SOPs) and various regulations also provide external structure and guidance. A powerful example of standard operating procedures are checklists. Checklists prompt and coordinate important cockpit tasks for each phase of flight (Degani and Weiner, 1990).

Besides heuristics and biases, there are other factors that can adversely affect decision making such as the speed-accuracy tradeoff, arousal, and preparation (Wickens and Flach, 1988). First, there is a speed-accuracy tradeoff between the time until an action is taken and time used to make a decision. In other words, people are more likely to make errors if pressured to respond quickly. The resulting danger of these errors is compounded in aviation because the stressful periods, when the possibility of speed-accuracy tradeoffs is greatest, can also be the least forgiving. Second, arousal might drive people to trade speed for accuracy. The arousal from cockpit aural alarms and alerts may create a sense of urgency to act that is likely to result in error. Lastly, even though reaction times will be faster for events people prepare for, preparation can also lead to inappropriate responses to unexpected events. Because people prepare to react to an expected event, they are likely to react the same way to unexpected events. (Wickens and Flach, 1988)

Wickens et al. (1988) used a computerized pilot decision-making simulator/trainer to analyze the decision-making components of low and high experienced pilots. Thirty-eight instrument rated pilots were divided into two groups based on reported hours of flight experience. Using 400 flight hours as a cutoff, the sample was divided approximately in half with all students in the low experience group (novice) and all instructors in the more experienced group (expert). All pilots took a cognitive test battery before testing. Then the pilots' performance was evaluated by the optimality and latency of their choices, and their rated confidence. The results showed little difference in judgment performance, although the experience pilots expressed more confidence in their decisions. However, the performance of the low experience group was partially predicted by some information processing tests but not for more experienced pilots.

External influences, such as the flight environment, have been found to account for "as much as half of the variability in pilots' problem solving behavior" (Casner, 1994, p. 580). In a study of ATC-cockpit transmissions, different flight clearances were analyzed along five dimensions. The five dimensions were: 1) Clearance type; 2) Clearance predictability; 3) Time constraints; 4) Average number of clearances per sector; and 5) Number of clearances issued at once. The types of clearances were headings, altitudes, and speeds or combinations of these and/or in association with a fix. Flight path management was conceptually broken down into three components: 1) Control (making small but continual adjustments to maintain a single heading, altitude, and airspeed); 2) Guidance (deciding what sequence of control responses is necessary to achieve a new target heading, altitude, or airspeed); and 3) Navigation (formulating, in advance, a sequence of headings, altitudes, and/or airspeeds that constitute an entire flight route) (Casner, 1994). For each flight path management task directed by ATC, the pilot can choose from three flight control resources to respond to; the manual autopilot, autopilot with flight director, or flight management computer (FMC). These three flight resources vary in the minimum amount of time required to start responding to a clearance. The FMC requires more time to enter data but is the preferred method for navigation tasks. In this study, pilots used the FMC in sectors where they could predict clearances and resorted to the autopilot in unpredictable sectors. Casner found that the pilots' choice of automation varied in relation to the predictability of an ATC sector.

From another angle, Layton, Smith and McCoy (1994) show three possible models of cooperative problem solving where the pilot can work with a computer in making decisions. The three system designs vary in the level of details provided by the computer. All three systems use tools supporting asking "what if" questions. Computer limitations, called "brittleness," are similar to weaknesses in human heuristics and biases. In essence, the computer makes unrealistic assumptions because it assigns all information equal values and only uses the information it is given.

Judgment

The constructs of judgment and decision making are often used interchangeably. However, research shows that judgment has a more global connotation. In aviation, judgment reflects how well decisions are made or the overall effectiveness and safety of a decision whereas decision making refers to the more technical aspects of how a decision is made. In general, performance research tests the efficacy of judgment training models and tries to discriminate between the judgment abilities of different pilots.

Jensen (1982) defines pilot judgment as: "1) The ability to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action, and to determine expected outcomes from each alternative. 2) The motivation to choose and authoritatively execute a *suitable* course of *action* within the time frame permitted by the situation, where: a) *Suitable* is an alternative consistent with societal norms; b) *Action* includes no action, some action, or action to seek more information" (p. 64).

Pilot judgment encompasses two different information processes (Jensen, 1982). The first is primarily perceptual-motor processes that involve highly learned perceptual processes that must be made in a short period of time or continuously such as assessing distance, altitude, speed, and clearance. The second is primarily a cognitive process of choosing a course of action from several alternatives where there are usually a number of complex considerations. The two types of judgment form a continuum of decision time versus cognitive complexity. Along this continuum, appropriate levels of time and analysis are allocated to a situation based on the circumstances

Buch (1984) outlines three factors influencing judgment in decision making; the pilot, the environment, and the aircraft. The pilot factor includes skill, knowledge, health, stress, fatigue, and other factors that might affect personal performance. The aircraft factor includes airworthiness, equipment, operating limitations, and any other factors that might affect aircraft performance. The environment factors are the mission, weather, terrain, ATC, or any outside information or conditions that affect the mission, pilot, or aircraft.

Stone, Babcock, and Edmunds (1985) reviewed Aviation Safety Reporting System (ASRS) reports for commonalties in pilot judgment error. The personal accounts of what happened were studied to determine why the decision-making failed. Out of 70 reports of 64 separate incidents, many different situational factors were found and no recurring problem stood out. Some problems were deliberate and not time-critical, but the majority involved time criticality or pressures. Limited information was also often a factor.

Most mishaps occur not from a single bad decision but from a chain of errors or poor judgment (Alkov, 1994). Poor judgment can be reduced by accurate information processing on the individual and crew levels. There are five steps to breaking a poor judgment chain: 1) Recognizing and admitting a poor decision; 2) Checking stress levels which can reduce a person's ability for good judgment; 3) Identifying the dangerous results of poor judgment and correcting them; 4) Being vigilant for other poor decisions because poor decisions create inaccurate information for other decisions; and 5) Reviewing the bad judgment to avoid similar faulty decisions (Alkov, 1994).

Jensen (1982) proposes that judgment can be evaluated by looking at an individual's willingness to follow rules, to resist peer pressure, to refuse to fly or turn around when situations deteriorate, to set limits based on personal capabilities, and to stick to personal limits on all flights regardless of the passenger's identity or importance of the mission. With these tendencies in mind, the evaluator considers: 1) Discriminative judgment: considering all relevant information and available alternatives, determining the relative importance of different information, and integrating relevant information efficiently before selecting a response and 2) Response selection tendencies: exhibiting any tendency to consider factors other than safety (such as self-esteem, adventure, social pressure, financial gain, or convenience) or semirelevant factors (financial gain or convenience) in situations where safety should have been the primary consideration (Jensen, 1982).

Buch (1984) lists the five hazardous thought patterns that are often examples of poor judgment in aviation literature because these attitudes are thought to be associated with aircraft accidents. These hazardous thought patterns are: 1) Anti-authority: resenting outside authority directing the pilot and disregarding rules and procedures; 2) External control: perceiving little control over life and attributes everything to luck or someone else's actions; 3) Impulsivity: acting quickly and on first thought; 4) Invulnerability: acting as though nothing bad can happen; and 5) Macho: trying to prove better than others, tending to be overconfident and attempt difficult tasks to gain admiration (Buch, 1984). He then proposes using tailored situational exercises requiring students to evaluate responses associated with hazardous attitudes.

There are several ways investigators typically evaluate pilot judgment by creating scenarios that require complex decisions. These methods include "Paper and pencil" methods, computers, simulators, and actual flights. These methods vary on the amount of realism and experimenter control.

Paper-and-pencil situational tests are what Motowildo, Dunnette, and Carter (1990) call low-fidelity simulations because they are a flatter, less-realistic simulation of job situations. Using subject-matter experts' critical incident descriptions and judgments of responses, They have constructed effective tests of future work behavior. They reason that high-fidelity simulations, using expensive technology to create close-to-actual work conditions, need to prove the added predictive validity.

A simple "paper and pencil" description with questions has been shown to discriminate between ATP and instrument rated pilots, private, and commercial pilots (Jensen, 1982). The scenario-driven test creates a divert situation where the pilot rates the relative importance of four factors in making the divert decision (Jensen, 1982). The four factors Jensen identified were air traffic control service (radar vs. no radar), the weather at possible destinations (ceiling of 1,000 or 500 feet), the time to fly to the airport (15 or 30 minutes) and the best approach facilities (ILS vs. ADF) at the alternate airport (Jensen, 1982,). A computer-generated scenario and/or flight simulator could create a more complex, realistic, and time-restricted situation where the pilot can exercise judgment. However, to further realism and judgment opportunities, a computer can use a set of stored algorithms to generate different responses to a student's questions.

Besides simulations, actual work situations can be used to test judgment. Embry-Riddle Aeronautical University devised a judgment training program and then tested its efficacy with volunteer student pilots in their flight school (Buch, 1984). The students were tested on short solo flights by observers putting them in judgment situations that seemed logically associated with the flight. The observers were ostensibly just observing and asking questions but actually evaluated the entire flight. The students who had been taught about judgment concepts exhibited significantly better judgment on 13 of 18 items. Unfortunately, testing judgment in the aircraft does not ensure consistent testing conditions.

Aeronautical Decisionmaking (ADM) is judgment training based on developments in cognitive psychology and is used as a cognitive model for research (Adams, 1993; Diehl, 1991b; O'Hare, 1992). ADM programs target pilot attitudes and behavior. Pilots are taught basic concepts of how errors develop and how to prevent errors. ADM can be diverse because it is often tailored to a specific aircraft and comes in many formats (manuals, lectures, flight training) and perspectives.

Research looks for reduced accident rates as indicators that ADM training works. The two methods of investigating ADM accident-prevention effectiveness are through experimental conditions and observational evidence. Diehl's (1991b) summary of six different experimental evaluations with low time general aviation pilots shows error rates decreasing between 8 and 46 %. Likewise, Bell Helicopters Texton Inc. (Diehl, 1991b) show 36-48% drop in operational accidents after implementing ADM. These figures show a dramatic decrease in controlled-setting errors and operational accidents which is a potential sign of validity and effectiveness of the ADM models.

Workload

Workload is defined as the relationship between the task and environmental demands and the human's ability to cope with these demands (Gopher & Braune, 1984). Cohen, Wherry, and Glenn (1996) emphasize that workload research tests the capacity limits of the operator rather than whether or not the operator can perform all of the assigned tasks. Hence, workload can be conceptualized as capacity for performance. The operation of modern flight systems requires pilots to rapidly process large amounts of complex information. Subsequently, this large demand may exceed the mental capacity of the pilot. Too much workload on one crew member is often associated with serious errors (Ruffel Smith, 1979).

Ordinarily, most pilots have plenty of spare capacity which may mask differences in performance potential (Svensson, Angelborg-Thanderz, & Sjoberg, 1993). Only under high workload situations may the ultimate differences in performance become evident. Nevertheless, every task hypothetically has a psychological and physiological cost associated with the task's workload (Svensson et al., 1993). This cost has been referred to as an investment of energy or resource consumption (Cohen et al., 1996).

The mental workload is composed of the mission requirements and conditions. The requirements include the number, difficulty, and sequencing of tasks. Likewise, the conditions include the number and significance of detrimental conditions. These conditions can be external (e.g. weather, equipment limitations and malfunctions, crew conflicts) and internal (e.g. fatigue, inexperience). "Both external and internal conditions impair the pilot's capability to process information, make decisions, and act. Accordingly, the pilot's mental workload is affected by both psychological and physiological factors" (Svensson et al., 1993, p. 986).

There are three general approaches to measuring workload. These approaches are: 1) Objective parameters of the task(s) or performance-based techniques, 2) Behavioral and physiological responses of the individual, and 3) Subjective ratings of the individual (Gopher & Braune, 1984; Svennsson et al. 1993; Wierwille & Eggemeier, 1993). Each of these approaches offer different perspectives and measurements.

Objective parameters of tasks can be used to estimate workload by the logic that the more number or difficulty of tasks will create more work. An advantage of this kind of objective measure is that operator performance can be measured easily, directly, and precisely (Hart & Hauser, 1987). In addition, the task-performance measures can focus on primary or secondary task performances (Wierwille & Eggemeier, 1993).

Performance-based workload techniques need to consider sensitivity, intrusion, diagnosticity, global sensitivity, transferability, and implementation requirements (Wierwille & Eggemeier, 1993). A sensitive measure distinguishes differences in workload. Any intrusion or part of the workload measuring technique that causes changes in the operator-system performance confounds results. Diagnosticity is the ability to identify the kind of workload, what causes the workload, and how the workload relates to the task (Wierwille & Eggemeier, 1993). On the other hand, global sensitivity is being able to measure changes in resource expenditure or factors that influence workload (Wierwille and Eggemeier, 1993). A measure's transferability is the extent that it can be applied to other situations. There are numerous practical implementation considerations such as required equipment to test and record, data collection procedures, and training of both the operator and subjects.

The multiple resource theory offers an explanation of how resources are used to perform workload tasks. This theory divides human capabilities into 7 discrete resources or resource channels: Visual, Auditory, Spatial, Verbal, Analytical, Manual, and Speech (Cohen et al., 1996). Multiple resource theory is the basis for the Workload Index (W/INDEX) model. In one study using the W/INDEX, pilot estimates of mission task workloads showed high inter-correlations and little meaningful discrimination (Cohen et al., 1996).

Hart and Hauser (1987) used communications performance as an index of significant sources of workload. Communications require that the pilot hears and understands messages as well as complies with any directed actions. Previous research found that communication tasks can be sensitive to varying amounts of workload. Another practical consideration is that communication behavior can be recorded easily. "Furthermore, there is a direct mapping between pilot response and measurable output that is independent of the system lags, vehicle dynamics, and operating characteristics that influence the performance measures available for other tasks" (Hart and Hauser, 1987, p. 403).

There is a large variety of behavioral and physiological responses that may be associated increased workload. Central nervous system measures such as electroencephalographic (EEG) activity and the event-related brain potential (ERP) have been used. Other physiological measures include heart rate, heart rate variability, pupillary dilation, and endrocine reactivity (Svennson et al., 1993). In a similar way, stress reactions have been measured by monitoring plasma cortisol, catecholamines, lactate, total protein, osmolality, total lipid, glucose, and haematocrit levels because these are associated with stressors in certain animal species (Sive and Hattingh, 1991). A problem with using these physiological variables is that each has a different function and is part of different reactions which are difficult to integrate into a single "stress index" (Sive and Hattingh, 1991). Also, the biological variables are interdependent in the system and may change in response to system changes elsewhere (Sive and Hattingh, 1991). Often, these physiological variables present very different and potentially misleading profiles.

Subjective measurements of workload rely on the performer's estimate of difficulties in the performance of a given task. The advantages are that these measures are easy to get and have a very high face validity (Gopher, 1984). Phenomenologically, if a subject describes a high workload, then the subject still perceives a high workload no matter what the behavioral and performance measures say. These subjective appraisals of workload have been shown to offer highly consistent profiles despite diversity of tasks and subjects in a study (Gopher and Braune, 1984). The consistency of these subjective measures indicates the existence of some kind of psychological attribute, "resource requirements," that can be measured and is related to characteristics of tasks. Some subjective measures of workload include the NASA Task Load Index (NASA-TLX), Cooper-Harper rating scale, bipolar rating techniques, and the Subjective Workload Assessment Technique (SWAT). The NASA-TLX measures aspects of mental, physical, and temporal demands along with performance, effort, and frustration level. SWAT is conjoint measurements of time load, mental effort load, and stress load which are translated into a workload index.

The different task measures, physiological measures, and subjective ratings are often combined in either a test and evaluation (T&E) environment or actual flight. "There is a general agreement that workload is a multidimensional construct. This implies that evaluations of several task-related and operator-related aspects of pilot's activities should be included in a rating scale to obtain the most accurate assessment of a pilot's experience" (Hart and Hauser, 1987, p. 403).

Svensson, Angelborg-Thanderz, and Sjoberg's (1993) structural linear causal model illustrates how workload factors may interrelate in the performance process. The model starts with the tasks or challenge factor (difficulty and risk). The working part of the model is the task performance which is modified by dual coping-process reactions of problem solving and emotional coping. The model ends with the specific and general performance outcomes.

In this model, the problem-solving process is characterized by commitment and activation. Activation and commitment indicate psychological "energy mobilization" which promotes efficient problem-solving, decision-making, and direct action (Svensson et al., 1993). In fact, activation (mental energy) may be the opposite of mental workload where mental energy is "the ability to regulate successful action in the face of obstacles such as fatigue and fear. Positive expectations, interest, and job motivation in general would be candidates for inclusion in such a measure" (p. 991). The predominant mood of problem-solving process is "active and alert" which affects performance positively. On the other hand, an emotion-coping process is "characterized by tension, effort, and adrenaline reactivity. Increased challenge results in increased tension which, in turn, increases effort and decreases activation" (p. 988). The predominant mood of emotion-coping process is "tense, under stress" which impacts performance negatively. Workload can be studied using "the variables included in or directly affected by the emotion coping process; i.e. tension, psychological and physiological effort, adrenaline reactivity, and activation (inverted) constitute the markers of the mental workload index (WI)" (p. 988). This model of workload has shown that challenge increases both problem-solving and emotion-coping where problem-solving improves performance and emotion-coping impairs performance. In a study with ground attack pilots, activation (problem-solving) ratings were higher and tension (emotion-coping) ratings were lower than norm data. This problem-solving dominance is assumed to last as long as the pilot's capacity is greater than the workload. But when the mission workload exceeds pilot capacity, the affects of emotion-coping becomes more evident.

Inflight measures of workload were taken on pilots flying 11 routine missions in a NASA aircraft used as an airborne observatory (Hart and Hauser, 1987). These measures included communications performance, subjective ratings, and heart rate, and were taken during seven flight segments: 1) Pre-flight briefing to taxi for takeoff; 2) Takeoff to flaps up; 3) Flaps up to start of astronomy recording; 4) Data recording to mission mid-point; 5) Mission mid-point to start of descent; 6) Start of descent to approach flaps; and 7) Approach flaps to landing roll. The three different measures demonstrated slightly different relationships with the various aspects of the flight. "The pilot and copilot ratings of workload, effort and stress were sensitive to variations in flight-related task demands across segments, but did not reflect specific differences in type of demands imposed on the pilot and copilot" (p. 408). "The rate of communications per minute of flight provided the most sensitive communications-related indicator of workload...significantly related to workload, stress, and effort ratings and to average heart rate across flight segments" (p. 408). Moreover, "the heart rate measure was able to discriminate among flight segments and between aircraft commanders and copilots" (p. 408).

Compared to flight measurements of workload, research tends to prefer test and evaluation environments because of the ability to create and standardize workload tasks and environments. Experimental tasks such as monitoring and reacting to six gauges while doing concurrent arithmetic tasks are used to simulate single and dual tasks (Humphrey & Kramer, 1994). The workload measurement also uses NASA-TLX subjective ratings and ERP physiological measurements. Performance takes into account time to react (RT) and accuracy data for the monitoring and arithmetic tasks. In Humphrey and Kramer's study, specific conditions on the different gauges directed simple keyboard responses, where some of the gauges presented information with high predictability (HP) and the others with low predictability (LP). Following each block of trails, workload was rated with the NASA-TLX. In this study, ERPs seemed to be sensitive to changes in workload and performance. The rating and performance data "suggested that increases in the level of perceived workload and effort are a function of increases in the level of difficulty from the concurrent tasks and from the HP to LP version of the monitoring task" (Humphrey and Kramer, 1994, p. 11).

Gopher and Braune (1984) describes six patterns of how the perception of workload is related to characteristics of tasks. They found that people have no difficulty rating workload despite "differences in stimulus and response modes, and the variability in mental operations and transformation requirements" . They also found that dual-task situations are rated higher than the sum of individual ratings. Third, "replications and practice had ordered effects on the rating, such that difficult tasks were decreased, whereas easy tasks remained largely unchanged" (p. 529). Fourth, "by constructing a psychophysical power function, it was possible to predict dual-task perceived loads from the derived scores of single tasks by applying a simple additive rule" (A psychophysical function assumes that the subjective measures of workload correspond to a proportionate investment of processing faculties in performing tasks). Fifth, "estimates of resource requirements correlated highly with an index based upon the processing characteristics of tasks" (The resource requirements refer to a psychological workload attribute). Lastly, the estimates of resource requirements had low correlation with measures of task performance.

In many ways, workload seems to have a similar conceptualization and impact as stress does on performance. Too little workload or stress can lead to boredom and complacency, whereas, too much workload or stress can be overwhelming. Both too little and too much workload and stress can degrade performance and be dangerous. Nevertheless, a moderate amount of workload or stress seems to be a psychophysical catalyst for good performance. Yerkes and Dodson delineated this principle as early as 1908.

Task and Resource Management

The time-limited, multi-tasks of aviation require strategic behavior that schedules tasks and resources effectively. The scheduler has the choice of when and how to do tasks and use resources. Often the task and resource management follow patterns. These patterns can be studied using scheduling theory from operations research. Scheduling theory provides a conceptual model of individual and group planning in complex human-machine settings. Using scheduling theory, normative optimal and satisfactory scheduling strategies can be identified and used as performance norms. The emphasis is on whether or not the scheduler can choose a suitable sequence of tasks and resources to reach the goal. (Dessouky, Moray, & Kijowski, 1995; Moray, Dessouky, Kijowski, & Adapathya, 1991)

Scheduling rules are how tasks are sequenced and resources are allocated to tasks. Some basic scheduling rules are "first come, first served" (FCFS) also known as "earliest ready (arrival) time" (ERT), "shortest processing time first" (SPT), and "earliest due date (deadline) first" (EDD). Often people use a combination of these rules such as FCFS, utilizing available resources for routine jobs as they arrive, while giving priority to jobs with tight due dates. On the other hand, schedulers may use SPT to maximize the quantity of jobs completed or EDD to minimize task lateness. It is important to note that some tasks, such as extending the gear before landing, cannot be late. (Dessouky et al., 1995; Moray et al., 1991)

The basic components of a scheduling problem are the number and characteristics of the different task configurations and resources. The characteristics of each component can be detailed on a complex level. For ease of illustration, each component is either single or multiple elements with unique time and use restrictions.

Task configurations refer to the number of subtasks and task parameters. The task configuration can be single or multiple subtask processes. A single subtask process might be monitoring a flight instrument. A multiple-subtask process, on the other hand, is a combination of subtasks such as the steps landing an airplane. The task parameters include ready times, processing times, and due dates as well as times required for preparation or switching from another process, priority, and whether the process can be interrupted. (Dessouky et al., 1995; Moray et al., 1991)

Resources can be classified as either durable or convertible; singular or multiple. In the case of flying, some durable resources are the crew and aircraft. Examples of convertible resources are fuel, sensory stimuli, sensory information, and the mission definition. The task-scheduling process uses durable resources to transform convertible resources into desired products and services and thereby achieve an objective, such as profit or service, in a given environment (Dessouky et al., 1995).

There are three levels of scheduling decisions for classifying objectives and resources: 1) strategic, 2) tactical, and 3) operational. Strategic decisions, the highest level, determine the overall mission objective (final outputs) and allocation of needed resources (initial inputs). This mission formulation leads to plans. Then tactical and operational decisions carry out the plans with intermediate transformation processes of lower-level inputs and outputs. Aviation objectives can be broken down as strategic objectives (missions) that are dependent on tactical phases of the mission (maneuvers and crew behavior) that are further reliant on operational factors (mental processes of situation assessment and decision making). Likewise, the same hierarchy of decisions are applied to resource management where the strategic level is composed of the overall organizational resources (e.g. flying squadron), the tactical level is lower level organizational resources (e.g. crew and aircraft), and the operational level is the lowest level of resources (e.g. pilot or mental resources such as attention, perception, and memory). (Dessouky et al., 1995; Moray et al., 1991)

Reviews of safety studies and simulator research reveal several task management tendencies. Roughly a quarter of accidents reviewed in one study involved task management errors and over half of these errors involved failure of appropriate task initiation or termination (Raby and Wickens, 1994). These task management and shedding errors have also been found to be associated with the loss of geographical orientation. In other studies using flight simulators and increased workload, pilots maintained performance on high-priority tasks while degrading or shedding performance of low-priority tasks. But pilots did not perform lower-priority, secondary tasks at lower workload periods, even when possible. Another simulator study showed that less effective decision-making crews tended to schedule activities later in flight than did the effective crews.

Raby and Wicken's (1994) simulator study of task management examined three categories of pilot abilities to prioritize and shed tasks. Specifically, the study concentrated on "*when* people chose to perform tasks and *how* they chose to adapt to high workload periods" (p. 235). The three categories of tasks included MUST, SHOULD or COULD BE DONE. A MUST task needed to be done before reaching the ground or it impacts flight safety (e.g. lowering the landing gear). A SHOULD task does not necessarily affect safety but has the potential to be dangerous for the pilot, ATC, or other aircraft if not accomplished before reaching the ground" (e.g. setting up for a missed approach). A COULD task does not affect anyone else and can be done at a later time (e.g. completing paperwork). Out of seven tasks, the higher-priority tasks were Map reading, Answering ATC, and Setting up instruments. The lower-priority tasks were Safety checks, Calling ATC, Paperwork, and an arbitrary task of Encoding altitude. The study showed a general pattern as workload was increased, higher-priority tasks were done more often and lower-priority tasks were done less often. The better performing pilots completed tasks earlier and also were more flexible in alternating between tasks (at least more frequently)". Likewise, lower performing pilots were characterized by taking too long on a particular task and waiting too long to initiate critical tasks.

CHAPTER VIII

AN INTEGRATED OPERATIONAL PERFORMANCE MODEL

An integrated performance model would combine human factors concepts with operational performance measures and variables, offering several possible advantages. First, each of these concepts, measures, and variables show new possibilities in performance research, yet each represent largely independent research efforts. Meaningful combinations of these concepts, measures, and variables may better explain and accentuate their interaction and influence on performance. Second, these meaningful combinations may offer an easier framework for pilots and researchers to understand performance. Lastly, the inclusion of operational criteria and predictors may represent more accurately both the time-critical multi-task and multi-resource operational environment as well as operational performance.

A systemic orientation allows a meaningful synthesis of the many factors and processes that impact performance of individuals and groups. Both the pilot and aircrew can be perceived as interdependent systems consisting of common factors. These factors can be called resources because each can influence the system's performance. The systemic relationships or processes represent how the resources are used. Effective performance depends on appropriate, coordinated use of each system's resources, both personal and interpersonal.

A similar proposition is the fusion between individually focused Aeronautical Decisionmaking (ADM) and group-focused Cockpit Resource Management (CRM). Diehl (1991) points out that the functional distinctions between ADM and CRM are disappearing because both are concerned with management of attention, crew, stress, mental attitude, and risk. Both ADM and CRM illustrate how cognitive, affective, and behavioral measures can be effectively integrated using process-oriented functions.

There is evidence that such a perspective is beneficial and works. Individual and group-related errors are the major causes of aviation accidents (Foushee, 1984). The proper management of personal and interpersonal resources can help reduce the number and gravity of accidents as well as potentially improve operational-success measures such as fuel use, equipment life, passenger comfort, coordination with users, on-time takeoffs and arrivals. There is experimental and observational evidence that resource management programs, like ADM and CRM, work.

Human factors and cognitive studies discuss fundamental processes that are not yet effectively operationalized in relation to pilot performance. The personal processing concepts include situational awareness, judgment, decision making, and workload management. The processes are basically information-process theories, many of which are still in the exploratory stages of development. Presently, information processing variables are operationalized only as very discrete tasks compared to the more global conceptual patterns that are hypothesized. However, "situational awareness" and "judgment" are common aviation jargon for more global concepts. These concepts capture the crucial aspects of flying but are difficult to empirically measure and evaluate. All of these concepts have been studied independently but there is considerable holistic interdependence when it comes to effectively and safely flying a mission. In essence, these measures constitute personal resource management which is the efficiency of resources use. Both on personal and interpersonal levels, these variables and their systemic relationships seem to describe a process of resource management.

The processes capture the personal and interpersonal dynamics of acquiring and acting upon information. Information-processing variables can be divided into intrapersonal dynamics within individuals and interpersonal dynamics between individuals. Personal processes include patterns of workload management, situational awareness, prioritization and task management, problem solving, decision making, judgment, and feedback. Interpersonal processes include the group patterns of leaders and team members. These processes are based on cognitive models and many of the processes such as situational awareness and workload management are applied to both individuals and groups. There are strong similarities in the dynamics of how these processes work on the individual and group level.

There is also an undeniable affective component to these processes. It is essential to understand and include affective factors, such as relevant personality traits and attitudes including stress-coping abilities, to the extent that they moderate resource management. These emotional factors can be conceptualized as basic continuums such as levels of confidence, motivation, and stress-coping. This conceptualization may require a complex consideration of emotional measures. For example, an internal locus of control could translate into strong motivation and confidence which enhances decision making and interactions. On the other hand, an external locus of control could contribute to poor stress-coping and thereby adversely impact decision making and relationships. The studies on experts, safety, crew resource management suggest influential personality factors exist but their connection to performance is difficult to measure. Resource management models may be able to integrate how common pilot personality characteristics such as competitiveness, achievement, dominance, self-sufficiency factor in to different levels of performance. Personality factors can be integrated to the extent that they facilitate or impede personal and group resource management performance.

Svensson, Angelborg-Thanderz, and Sjoberg's (1993) workload model offers a simplified conceptualization of how affective components might affect performance. In this model, problem-solving dominates emotional-coping until the mission demands reach an individual's mental capacity. Then emotional-coping is more evident and impairs performance because the tension and stress of emotional-coping overshadow the active and alert problem-solving.

A more comprehensive, accurate, and functional model is possible in which affective components can either bolster or undermine cognitive performance. Diehl (1991) offers a schematic of the interrelationships between six aeronautical factors (abilities, motivation, knowledge, procedures, perceptual-motor skills, and decisional judgment) where the three basic types of errors (procedural, perceptual-motor, and decisional) interfere in normal feedback loops. All of these factors interact with the situation which is composed of the tasks and environment as well as the aircraft and crew.

Likewise, the relevant situational elements can be factored into a resource management model. The Hughes Training CRM workbook outlines main categories of elements in a risk management model. These elements are the: Aircraft, Environment, Situation, Operation, and Personnel (AESOP). Each element includes many variables and the situation element is a reminder that risk is synergistic so the total risk is greater than the sum of the individual risks.

The processes can be used to develop predictors based on a personal and interpersonal resource management model. These process-oriented pilot performance predictor variables could integrate individual and group processing concepts, expert opinions and examples, personality and stress research, and safety research into viable performance measures. These process performance measures could be compiled in to personal and group resource management profiles. Profiles of personal and group resource management could then be developed for selection and training purposes. These profiles would be behaviorally-anchored scales with non-clinical descriptions, benefiting both pilots and scientists.

In this way, the profiles could be less threatening for and easier to understand and be used by pilots. The process profiles would put into a common language what was known on a more intuitive level by expert pilots who evaluate pilot performance and recommend upgrades or further training. In addition, the scales could even help evaluators rate pilots and give feedback. It is valuable for a pilot to be aware of his/her tendencies and limits and to have concrete descriptions of these tendencies for future training. The process descriptions would reveal the tendencies or patterns that are causing the poor performance. By focusing on patterns, the critical commonalties of performance are not lost in the details and valuable connections are realized. The overall systemic perspective should help pilots conceptualize and discuss how various behaviors and/or tasks interact. For example, a pilot may be struggling in seemingly unrelated areas, and connections between these areas could be identified by a systemic perspective. On the other hand, there are also many ways to complete the same task and a systemic perspective can highlight the effective elements of each way. Lastly, performance outcome may misrepresent how a task is completed and the systemic tells more of how performance might improve or be affected by different circumstances (Chidester et al., 1990). For example, short cuts or incomplete planning may often work with ideal or lucky conditions but do not reflect consistent performance. Instead of focusing on the outcome which can often be deceptive, the focus is on the patterns.

For scientists, the resource management systemic could meaningfully unite and clarify the inputs and outputs of flying. The cognitive, affective, and behavioral inputs are the tools which are only as effective as how they are used. Essentially, processes could evaluate how a person's tools (resources) are used. Another advantage of looking at systemic relationships is capturing and integrating the importance of seemingly minor tasks like planning and communicating. Lastly, researchers may be able to compare similar systemic relationships across different aviation environments and tasks.

The personal and interpersonal resource management model can set the stage for useful and safe computer-assistance. Computer systems will be increasingly part of aircraft systems. The goal of computers is to help aircrew deal with the complex aircraft systems, tasks, and environments; however, these computer systems will help only as far as they can fill needed gaps in personal and interpersonal resources. Careful attention must also be given to the fundamental limits and quirks of human systems. Without knowledge of the characteristics of human information-processing, computers could work at cross purposes with human systems. Just as pilots need to be wary of the limitations of highly proficient but naive computer systems that act like "third pilots," the computer systems need to take into account the human personal and interpersonal resource management tendencies.

Previous UPT measures are probably not sensitive to these factors. The value of information processing may be underestimated because it may be closely related to general cognitive abilities and is not adequately tested by UPT tasks and environments. Pilot training performance is due to "not much more than *g*" (Olea and Ree, 1994). Information processing is probably a more-specific, higher-functioning cognitive ability that is overshadowed by general cognitive ability performing the basic flying tasks of UPT. This would explain why the importance of information processing may not be relevant until the later, more demanding stages of pilot training.

The various UPT performance measures do not adequately represent the complex depth of operational situations and breadth of operational responsibilities. Measures of performance with more complex environments and tasks provide better reflections of effectiveness and safety. In fact, the impact of safe attitudes and practices are more evident and critical in complex situations where there are more opportunities to make mistakes. A personal and interpersonal resource management model details how pilots safely and effectively approach these complex situations and responsibilities. The effective and safe management of complex situations and responsibilities are the fundamentals of operational success and pilot performance.

The limited complexity of UPT does not fully represent operational demands, predict operational success, measure operational performance, or show that students can transfer training lessons to operational situations. UPT performance measures are not as complex as operational situations and responsibilities. Pilot training success criteria focus primarily on limited effectiveness measures in a structured environment. These criteria help predict who succeeds at pilot training but do not necessarily equate to operational success, which is the reason for selecting and training pilots. Since pilot training tasks, environments, and responsibilities are highly controlled, there may be a limit to measuring the ability to apply knowledge and skills to more unique and complex operational situations. The ability to deal with operational complexity may not be adequately evaluated in pilot training. Likewise, a student's ability to apply knowledge from training experiences to complex real-world settings, called "situated cognition," is not established and may be limited (Campbell and Lison, 1995).

Operational missions involve more complex tasks and environments than UPT. In operations, flying the plane is often a secondary task to being a weapon's platform and communications center compared with the UPT primary task of flying. Operational tasks and environments are also more variable than the canned, ritualized, and familiar UPT task profiles and environment.

The ability to fly is only one of many operational criteria used for selection and advancement. For example, consider the hiring dilemma of Air National Guard (ANG)/Reserve units and commercial airlines when selecting from a pool of qualified pilots. All of these pilot applicants have demonstrated technical flying skills. The variation in these pilots is due more to other abilities such as being able to deal with complex tasks and environments. In addition, organizations want to hire pilots who can work with crews, manage flying missions and contribute to the flying organization on the ground (Hedge et al ,1994). These are pilots who can handle the complex situations (tasks and environments) as well as the complex responsibilities (instructor, evaluator, leader, manager, crew member, organizational member) safely and effectively.

A performance profile of personal and interpersonal resource management is one way of conceptualizing safe and effective management of complex situations and responsibilities. Operational performance could be measured and scaled based on behavioral examples of how pilots manage their resources. In this way, criterion measures could shift from static measures such as pass/fail, grades, and time to complete phases of training to the dynamic, underlying processes critical to being an effective and safe pilot.

CHAPTER IX

PERSONAL AND INTERPERSONAL RESOURCE MANAGEMENT PERSPECTIVE

A combination of human factors systemic concepts and operational variables may be able to integrate cognitive, affective, and behavioral factors affecting operational performance. However, the independent development of each of these concepts, measures, and variables is complicated and the proposed holistic approach to studying performance would be even more complicated. Therefore, the following potential criterion measures, predictor variables, and research methods offer only some beginning ideas of how human factors and operational variables can be integrated in a personal and interpersonal resource management perspective.

Potential Criterion Measures

Operational pilot performance is the most valid criterion. Therefore, the performance measures need to represent the critical aspects of both effective and safe operations as well as discern different degrees of performance. The two types of potential criteria are a dichotomous measure of succeeding or not succeeding operationally and a performance grading scale based on behavioral examples of succeeding or not succeeding operationally.

The first type of potential criterion is an overall indicator of successful or unsuccessful operational performance. Successful operational performance can be identified as achieving higher positions and qualifications. Most aircraft assignments have a methodical progression from basic pilot to instructor to evaluator positions along with special qualifications. Upgrades are based on knowledge, skills, and judgment; not all pilots upgrade to instructor and not all instructors upgrade to evaluator positions. To qualify as successful performance, a pilot's upgrade level should be at a minimum instructor with additional mission ratings depending on the type of aircraft and mission. The top performers could be further limited based on being selected for aerial competitions like Airlift Rodeo and Red Flag, or special programs like air weapons school.

In addition, another sign of operational success is being identified by other pilots. Pilots who fly together on a regular basis may be best qualified to select which pilots are best at the operational flying mission. The members, staffs, and commanders of flying units can select their top performers through anonymous ballots or discussions.

On the other hand, unsuccessful operational performance can also be used as a measure of performance. Unsuccessful performance can be defined as being involved in an aviation mishap due to pilot error or not passing an operational checkride.

The second type of performance criterion is behavioral examples of superior and poor performance. These examples of operational performance can be organized as critical performance dimensions with behaviorally anchored grading scales for each dimension. Moreover, these behavioral examples can be collected from the experience and insight of experienced pilots as well as safety reports.

Top performers can provide performance examples indirectly as critical incident descriptions or directly as what they think are examples of effective/ineffective and safe/unsafe performance. Most of the top performers should be evaluators because their job is to set the standards, evaluate the performance of other pilots, and make recommendations about further training and upgrades. After all, it is evaluator's grades and rankings of UPT students which provide present UPT criteria.

Safety reports are another source of operational performance examples. These studies can be systemically analyzed for critical components of behavior that contributed to a worsening situation as well as behaviors that (would have) corrected a dangerous situation. The underlying dangerous and constructive behavior patterns can be used as behavioral anchors for a grading scale.

The criteria should represent both effectiveness and safety which are the twin pillars of operational performance. Both are critical to consistent performance, although effectiveness is the more prominent aspect. Effectiveness is producing the desired affect where the emphasis is on getting the mission done. On the other hand, safety may run counter to the mission and require reduced effectiveness. Sidestepping some safety considerations can increase operational effectiveness, repeatedly without a price.

Only rarely can safety become a factor in performance, but the cost can be much greater than an incomplete mission.

The safety aspect of performance is probably underestimated and neglected. There is so much redundancy and backups in modern crew aircraft that unsafe shortcuts and omissions are often corrected by crewmember intervention or go uneventfully unnoticed. The outcomes of these situations can be deceiving because poor individual performance can be masked by the interventions of other crew members and fortuitous circumstances. In this way, a pilot who takes short cuts may appear to be proficient and even more efficient than others while actually leaving out critical aspects of performance. Such a pilot may intentionally or unintentionally skip over important considerations and tasks which may rarely affect performance negatively, but are crucial when it does.

Potential Predictor Variables

The potential performance predictors are systemic variables that are unified by a resource management perspective of how information is processed and used to sequence tasks and accomplish objectives. These variables can be viewed on a personal and interpersonal level. On a personal level, the predictor variables are the human factors and ADM systemic variables: situational awareness, decision making, judgment, workload and task and resource management. Likewise, on an interpersonal level, the predictor variables include CRM concepts of leadership, communication, and adherence to standard operating procedures. In addition, on both the personal and interpersonal levels, stress coping and personality patterns are integrated according to how they affect resource management.

The personal predictor variables focus on how the pilot uses available internal resources (mental faculties) to assimilate and act upon available information. This management process is essentially situational awareness, decision making, judgment, workload and task and resource management using the mental resources of attention, STM, and LTM. These systemic predictor variables measure cognitive patterns such as favored production rules, heuristics and biases for different situational contexts. Likewise, measures could track how long until a pilot recognizes a problem, responds to a problem, or resolves a problem as well as the accuracy of problem analysis and the chosen response.

The interpersonal predictor variables represent how the pilot utilizes outside resources, especially other aircrew members. These variables provide qualitative measures of critical crew interactions including leadership, communication, and adherence to procedures and regulations. These qualitative measures can be oriented towards assisting or impeding interpersonal resource management.

In addition, a predictor that moderates both personal and interpersonal resource management is stress coping. Because performance degrades under stress, individual stress coping abilities and tendencies are important predictors. Most pilots can handle routine stressors. The best pilots are the ones who can handle unusual or extreme stressors. Therefore, some predictors of performance are how a pilot perceives and handles stress as well as his/her stress coping limits.

Likewise, since people tend to regress to dominant responses under stress, these dominant responses are markers of how individuals might perform habitually in critical situations. Bob Stephenson (personal communication, July 1, 1996), director of crew resource management training at Flight Safety International, describes how pilots can "display" and "talk" good attitudes in academic training but will fall back to more natural behaviors under stress. Likewise, the pilot often projects a different personality on personality inventories because of a similar self-report bias, termed impression management. In addition, these dominant reactions are difficult to measure because of the canned nature of many training and evaluation flights. Measuring ordinary and dominant behavioral responses to controlled flight-like stressors would enable researchers to evaluate hard-to-get-at characteristics of pilots.

There are also personality patterns that may enhance or limit personal and interpersonal resource management. For personal resource management, some possibly applicable measures are confidence, aggressiveness, compulsivity, field independence, locus of control, and defense mechanisms. Likewise, there are personality patterns that may affect interpersonal resource management such as instrumentality, expressiveness, social confidence, assertiveness, and extroversion. There is previous evidence that many of these personality patterns exist in pilot samples or have correlated with performance. The relationship of these personality patterns with performance may be better explained as moderating resource management which is largely cognitive information processing.

The importance of personality variables may also be more related to safe as well as effective resource management. Personality studies suggest that many pilots are very comfortable with risk and are achievement-oriented, competitive, and aggressive. These personality patterns, combined with an affinity to fly and face challenges, means that pilots are very "go-oriented". These characteristics are both emotional assets and liabilities. There needs to be internal and external factors to regulate the drive to push personal and environmental limits and meet challenges. Some examples of external factors are aviation regulations and working in teams where aircrew monitor each other. However, the last lines of defense are internal personality factors because there are not rules and monitors for every situation, especially when the pilot is the aircraft commander or sole pilot on a mission. These personality factors are often referred to as the professional qualities of judgment, responsibility and accountability. The opposite of safe qualities are the previously cited five hazardous thought patterns of anti-authority, impulsivity, invulnerability, macho, and resignation.

Potential Research Directions

A resource management perspective views operational performance as largely cognitive information-processing moderated by stress coping and personality patterns. Various research settings, criteria and predictors, and research methods can be combined using this resource management framework. Using advances in computers and simulator technology, researchers can represent complex task and resource operational demands and measure the corresponding abilities. However, these research settings vary in realism and experimental control. Likewise, the variables and their measurement techniques differ according to the research setting as well as the available technology and research methods. For example, comprehensive performance criteria are possible through critical incident techniques and more relevant predictors are available through measures that track cognitive processing and focus on more occupational, less clinically oriented personality factors. Finally, longitudinal studies provide a way to link these predictors to the operational performance criteria (for example, N-EFS; King and Flynn, 1995).

There are several potential research settings used to test resource management decisions in different situations. The research settings include paper-and-pencil tests, computer tests, simulator missions, and actual flights. They vary from abstract to realistic presentations of operational demands and from more to less experimental control. Accordingly, each research setting offers unique advantages and limitations. The goal is to find the best way of isolating different types of information processing abilities and tendencies as well as studying the influence of personality factors and stress coping.

The paper-and-pencil test offers the most experimental control and the most abstract setting. The standard format is decisional scenarios followed by multiple choice questions. These tests are easily standardized and administered as well as inexpensive. However, the tests are an abstract representation of a problem scenario because the situation description and clues are limited to a verbal description and possibly pictures. Likewise, since potential solutions are provided as multiple choice answers, the answers may influence the pilot's search for a conceptualization of the problem and/or the solution. The options are already provided and limited. In addition, there may be some options the individual would not have thought about as well as not providing the option the individual would have exercised. In actuality, the individual thinks of his/her own possibilities and may select the first solution that comes to mind. Moreover, even the wording of multiple choice questions may affect the decision process yet this wording inevitably does not represent the uniqueness of individual's perceptions.

On the other hand, computers and simulators offer more control and versatility of testing environments and tasks as well as less abstraction. The added control of outside influences helps standardize tests, impose more realistic and complex tasks, study all variables with more discrimination, and possibly avoid requirements for highly specialized technical support (Flynn et al., 1994). Moreover, the flexibility allows more diverse situations and measures. In addition, the increased realism enables the pilot to act as if s/he were actually in the scenario.

The basic tradeoff between computers and simulators is the added control of what the pilot sees and hears with computer tests versus the increased realism of a simulator mission. Overall, computers seem more capable of isolating specific information processing skills where simulators are better at providing realistic conditions and situations as well as an environment to study group interactions, personality patterns, and stress coping. However, besides this tradeoff, both computers and simulators provide many different approaches to measuring resource management abilities.

Giffin and Rockwell (1984) demonstrated how computer technology can track information processing. Critical in-flight events were simulated using a computer mock-up instrument-panel display of a Piper Cherokee Arrow. The pilots used the touch-sensitive CRT computer screen to ask for information on the instrument panel, interior conditions, exterior conditions. and Air Traffic Control. A complete time history of all data inquiries revealed each pilot's information-seeking strategy and possible assumptions.

In addition, instructional computer programs offer ways to teach skills while assessing learning and other information-processing abilities. Benton, Corriveau, Koonce, and Tirre's (1992) Basic Flight Tutoring System (BFITS) uses a computer program to teach students how to fly in progressive instructional modules. Many dimensions of the student's performance are monitored to ensure that they do not progress until reaching a satisfactory skill level in each module. The logged data includes response latency, correct responses, incorrect responses, incorrect response specifics, all scores, as well as a comprehensive list of aircraft configurations, control, and performance information. This recorded data can also potentially be used for assessing resource management abilities and tendencies.

Likewise, LaJoie and Lesgold (1992) use a computer program to instruct and assess problem solving abilities. The program, called Sherlock, records how people troubleshoot problems by tracking their analysis and numbers of hints required to solve a problem. The results focus on patterns of student performance rather than simply the performance outcomes.

On the other hand, simulators offer the ability to work in more realistic environments including lifelike motion and aircraft responses as well as aircrew interactions. Successful simulation scenarios have at least five essential elements including: realism, ambiguous problems, ongoing consequences, complicating factors, and a requirement for full crew involvement(Chidester et al., 1990).

Wickens et al, (1988) outline desirable ways to structure simulated scenarios. The structure appears to be limitless to the user while actually providing a constrained formal structure for research. In addition, the structure provides a pattern of deteriorating circumstances that often characterizes aircraft mishaps. In this way, problems are not the result of one poor decision or technical malfunction but rather a result of "several concatenated events opening successive 'gates' to an accident" (p. 19). The structure fills the dual roles of keeping the simulated flight on a mission profile while also allowing digressions into successively less optimal scenarios" (p. 19). This structure incorporates "core" scenarios that are parts of an optimal mission and "side" scenarios which are generally less favorable and become more probable as decisions are less optimal.

The most realistic but least controlled research setting is actual flying. Flying missions have been used successfully in various ways. Hart (1987) showed how communications patterns could be coupled with other measures to study workload. Likewise, Buch (1984) studied student pilot judgment by observing their responses to 18 judgment items on cross-country flights.

Overall, there are many different types of decisional situations available depending on time, money, personnel, equipment, and other available resources. According to Bob Stephenson (personal communication, 1996), simulated situations can be as simple as a group of pilots role-playing a scenario in a classroom. On the other hand complicated simulations could include "synthetic crew members" to ensure more standardization of the simulated environment. However, the opportunity cost of increasing reliability through standardizing scenarios is decreasing the predictive validity of operational performance.

In general, all simulated situations would be designed so that pilots would need to use minimal outside or special knowledge and maximum levels of awareness and judgment. In fact, simulated problems could be designed that require no knowledge about aviation. These situations would be predicaments where there is no one right answer but only a choice between undesirable alternatives. The focus is on *how* the decision is made instead of *what* decision is made.

Task measurements can follow precedents from previous research. Wickens et al. (1988) used seven performance variables including: 1) decision choice, 2) optimality, 3) decision time (latency), 4) decision confidence, 5) problem detection, 6) problem study time, and 7) mean reading speed. Raby and Wickens (1994) add that the embedded secondary-task performance should measure whether (or how often) such tasks are performed and not the latency of their performance, once initiated. Research can investigate scheduling routines to observe how tasks are initiated and sequenced normally as well as how people deal with overload.

Algorithms can also contribute realistic evolution of tasks and scenarios. For example, some problems have time constraints and will get worse if corrective action is not taken within a reasonable amount of time. Algorithms can trigger default conditions when a pilot does not intervene in time. Boolean logic can be used to coordinate delayed effects of decisions on subsequent scenarios. These programs can also keep track of the scenarios and research measures for post-mission analysis. (Wickens et al., 1988)

Performance criteria for simulated missions can be developed by questioning and testing top performers for common qualities. The top performers could be asked directly or indirectly what are the characteristics of top performers. Likewise, the experts could be tested with a battery of standard psychological tests or evaluated in simulated situations by other experts.

A fruitful indirect method is asking pilots to describe critical incidents that illustrated the difference between average and superior airmanship (Hanson et al., 1996). Then these behavioral descriptions would be "retranslated" by other pilots to confirm and distill the important aspects (Hanson et al., 1996). The critical incident technique (Flanagan, 1954) is an effective way to organize and operationalize observations in a form that can be tested in controlled conditions. This technique is a way to capitalize on the insight, experiences, and opinions of experts and top performers. In fact, the technique requires incidents from qualified observers. In the same way, experts could be asked to evaluate crew member actions in accident and near-accident safety reports.

The dimensions of pilot performance generated by critical incident and other techniques can help create a reliable, valid, and comprehensive performance rating scale. The description of each performance dimension could be expanded into effective and ineffective behavioral examples. These examples can be further refined into behavioral anchors for a grading scale of each performance dimension. A standardized grading scale could reduce some of the subjective grading differences noted of pilot training instructors and provide a common ground for evaluating performance. For example, CRM uses the Line LOFT Worksheet grading scale that provides explicit classifications of 14 aspects of crew performance

(Helmreich et al., 1990). These aspects include Advocacy, Abnormal management, Briefings, Communications, Conflict management, Critique, Decision making, Distractions, Group concern, Inquiry, Preparation and planning, Proficiency, Task concern, and Vigilance. These or similar items could be applied to individual performance as well. Most importantly, it is possible to create specific and simple standardized performance criteria by organizing and condensing the perceptions of operational pilots.

These criteria could be applied by groups of expert raters to achieve even more reliable and valid results. These experts could be both pilot and researchers/psychologists observing screening evaluations, checkrides, and missions. Using a team of pilot observers or psychologists can establish inter-rater reliabilities (Chidester et al., 1990). Likewise, both psychologists and pilots offer strengths as raters. Psychologists can notice subtler psychological markers. In Germany, for example, psychologists are part of an evaluation team that observes screening evaluations (Gnan, Flynn, and King, 1995). On the other hand, evaluator pilots can be kept blind to the purpose of the study, reducing rater-bias potential, and perceive events from a more operational perspective.

A resource management paradigm incorporates more specific as well as broad ranging cognitive abilities, stress coping, and personality predictors. Even though resource management is largely cognitive information processing, both stress coping and personality may moderate performance depending on the situational context.

Cognitive tests could be designed to measure processing abilities and aptitudes. These processing variables could be measured by creating multi-dimensional computer tests with interdependent simple and complex situations as well as variable task and resource constraints. In this way, time could be another variable requiring the sequencing and prioritization of multiple sources of information and multitasks. The tests would vary time, tasks, and available information to isolate specific process variables. Moreover, the computer could selectively give information and ask questions in sequences to test various processing abilities not possible with speed and power tests. Complex scenarios could even use several levels of if/then structure where the content of the next question is based on the previous question's answer. These sequential decisional situations could link together like a flight profile. Overall, situations and choices could be structured to represent specific processes and the tests would get progressively more difficult following the spiral omnibus pattern (Anastasi, 1988). Some of the cognitive measures might be how memory, attention, and critical thinking interact in situational awareness, decision-making, judgment, workload management, and task and resource management.

In addition, stress coping will affect performance under certain situations. Stress coping represents both cognitive and affective ways of dealing with perceived difficulties. Simulated problems could take into account how pilots deal with unexpected situations and when they revert back to primary or dominant modes of behavior. Different levels of stress could be imposed to evaluate workload management and dominant reactions. The dominant reactions are critical because most pilots can perform the regular job well but their performance will vary under stress because they will revert to dominant reactions. Stress coping could be measured as information-processing styles under stress, optimal stress levels, and the difference between the subjective perception and the physiological signs.

In the same way as stress-coping, personality patterns can target affective patterns relevant to performance in certain situations. In fact, one possibility is measuring personality patterns concurrently with cognitive testing. The combination of testing pressure and masked personality testing may eliminate some self-report bias as well as give personality patterns during demands similar to operational situations. In addition, global, nonclinical inventories offer more relevant personality profiles such as the NEO-PI-R (Costa and McCrae, 1992).

Longitudinal studies offer a way to study how resource management predictors relate to operational criteria. These longitudinal studies could originate during pilot training screening to different career points. In this way, relevant measurements could be connected to the pilots who are more and less successful operationally. Both the military and airlines do extensive screening so setting up such an archive should be possible. Higher performing groups could be compared with lower performing groups and these groups could also be compared to the general pilot population Likewise, longitudinal differences within groups may reveal important considerations about changes in resource management profiles for different groups of pilots. Possible confounds are different training environments and upgrade opportunities. Longer periods and broad samples may allow these differences to level out.

Operational criteria could be used in longitudinal studies beginning before UPT and spanning at least as long as the active duty service commitment of pilots in the Air Force. Personal and interpersonal resource management profiles could be measured to establish a baseline before UPT and then at regular intervals to monitor changes. Pilots who demonstrate consistent superior or poor performance could be examined in relation to these process measures of how they handle situations.

CONCLUSION

The ultimate goal of pilot performance research is to predict operational performance. Two difficult aspects of this research are finding operational criteria and relevant predictors. Overall, the conceptualization of these variables requires more relevant and broader measures of operational performance. More valid and reliable measures may be possible by integrating operational variables based on systemic frameworks. The combination of systemic and operational variables depict the complex interactions of multiple job factors and performance variables in operations.

Complexity is a major characteristic of the information processing and resource management tasks in aviation. There are many sequenced tasks to accomplish by coordinating various resources on several levels. The two basic levels are the individual pilot and crew. The performance dynamics on these levels can be conceptualized as personal and interpersonal resource management.

A personal and interpersonal resource management perspective shifts the research focus to a more holistic and positive view of how a person organizes and approaches tasks with available resources. ADM and CRM are established and successful examples of systemic perspectives. These models also incorporate or can be fortified by the development of human factors systemic concepts such as situational awareness, decision making, judgment, workload, and tasks and resource management.

Complexity also characterizes how many variables affect operational performance. Systemic models facilitate an operationally meaningful combination of cognitive, affective, and behavioral variables. This meaningful combination describes the complicated interdependence between variables and, thereby, better explains how the variables affect performance. In general, the relevant performance variables appear to be largely cognitive and psychomotor. However, affective variables can be included to the extent that they moderate cognitive and psychomotor performance.

In addition, a systemic and operational focus encourages a broader operational perspective of both performance criteria and predictors. Aviation involves much more than just flying the aircraft. There are other operational roles and responsibilities that are also important performance dimensions. Likewise, the inclusion of wider-ranging performance criteria illustrates how other predictors affect performance. The operational relevance of these variables is explained by studies using pilot perceptions of performance, occupationally oriented personality measures, and safety research on personality, stress coping, and error types.

There are many issues to resolve in developing the proposed models, variables, and testing procedures. The integration of the various mental models requires knowledge of the latest developments in cognitive and affective studies as well as the complicated human factors models. There are also questions about determining what personality dimensions and processing abilities are fairly rigid and what are more amenable to training. The questions should be answered as the different models naturally move to a more integrated and operational focus.

References

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. Human Factors, 37 (1), 85-104.

Adams, R. J. (1993). How expert pilots think: Cognitive processes in expert decision making (Tech . Rep. No. DOT/FAA/RD-93/9). Jupiter, FL: Advanced Aviation Concepts.

Adams, R. R. & Jones, D. R. (1987). The healthy motivation to fly: No psychiatric diagnosis. Aviation, Space, and Environmental Medicine, 58, 350-354.

Alkov, R. A. (1987). When the balloon goes up, so does the mishap rate. Approach, 33 (1), 2-6.

Alkov, R. A. (1994). Enhancing safety with aircrew coordination training. Ergonomics in Design, April, 13-18.

Alkov, R. A. (1996). Human Factors in Aviation. Kirtland AFB, NM: Southern California Safety Institute.

Alkov, R. A., Gaynor, J. A. & Borowsky, M. S. (1985). Pilot error as a symptom of inadequate stress coping. Aviation, Space, and Environmental medicine, 56, 244-247.

Anastasi, A. (1988). Psychological Testing (6th ed.). New York: Macmillan Publishing Company.

Baker, S. P. (1995b). Putting "human error" into perspective. Aviation, Space, and Environmental Medicine, 66 (6), 521.

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

Baker, S. P., Li G., Lamb, M. W., & Warner, M. (1995a). Pilots involved in multiple crashes: "Accident proneness" revisited. Aviation, Space, and Environmental Medicine, 66, 6-10.

Bale, R. M., Rickus, G. M. & Ambler, R. K. (1973). Prediction of advanced level aviation performance criteria from early training and selection variables. Journal of Applied Psychology, 58 (3), 347-350.

Banich, M. T., Stokes, A., & Elledge, V. C. (1989). Neuropsychological screening of aviators: A review. Aviation, Space, and Environmental Medicine, 60, 361-366.

Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-26.

Baumrind, D. (1970). Socialization and instrumental competence in young children. Young Children, 26 (2), 104-119.

Benton, C. J., Corriveau, P., Koonce, J. M., & Tirre, W. C. (1992). Development of the Basic Flight Tutoring System (BFITS) (Tech. Rep. No. AL-TP-1991-0060). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Blower, D. J. (1992). Using constraint satisfaction networks to study aircrew selection for advanced cockpits (NAMRL Special Rep. No. 92-1). Pensacola, FL: US Naval Aviation Medical Center.

Boer, L. C. & Castelijns, M. T. (1991). The PILOT test as a predictor of pilot aptitude. In M. Dentan and P. Lardennois (Eds.), Human factors for pilots: Proceedings of the XIX Western European Association for Aviation Psychologists (WEAAP) conference (pp. 1-12). Aldershot, England: Avebury Technical.

Bordelon, V. P. & Kantor, J. E. (1986). Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies (Tech . Rep. No. AFHRL-TR-86-4). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Buch G. (1984). An investigation of the effectiveness of pilot judgment training. Human Factors, 26 (5), 557-564.

Butcher, J. N. (1994). Psychological assessment of airline pilot applicants with the MMPI-2. Journal of Personality Assessment, 62 (1), 31-44.

Butcher, J. N., Jeffrey, T., Cayton, T. G., Colligan, S., DeVore, J. R., & Minegawa (1990). A study of active duty military personnel with the MMPI-2. Military Psychology, 2 (1), 47-61.

Campbell, J. O. & Lison, C. A. (1995). New technologies for assessment and learning. College Student Journal, 29 (1), 26-29.

Carretta, T. R. & Ree, M. J. (1994). Pilot-candidate selection method: Sources of validity. The International Journal of Aviation Psychology, 4 (2), 103-117.

Carretta, T. R. & Siem, F. M. (1988). Personality, attitudes, and pilot training performance: Final Analysis (Tech . Pap. No. AFHRL-TP-88-23). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Casner, S. M. (1994). Understanding the determinants of problem-solving behavior in a complex environment. Human Factors, 36 (4), 580-596.

Chambers, A. B. & Nagel, D. C. (1985). Pilots of the future: Human or computer? Computer, 18 (11), 74-87.

Chidester, T. R., Kanki, B. G., Foushee, H. C., Dickinson, C. L. & Bowles, S. V. (1990). Personality factors in flight operations: Volume I. Leader characteristics and crew performance in a full-mission air transport simulation (NASA Technical Memorandum 102259). Moffett Field, CA: NASA-Ames Research Center.

Chidester, T. R., Helmreich, R. L., Gregorich, S. E., & Geis, C. E. (1991). Pilot personality and crew coordination: Implications for training and selection. The International Journal of Aviation Psychology, 1 (1), 25-44.

Cohen, D., Wherry, R. J., & Glenn, F. (1996). Analysis of workload predictions generated by multiple resource theory. Aviation, Space, and Environmental Medicine, 67, 139-45.

Corsini, R. J. (Ed.) (1994). Encyclopedia of Psychology. New York: John Wiley & Sons.

Costa, P. T. & McCrae, R. R. (1987). Validation of Five-Factor Model of personality across instruments and observers. Journal of Personality and Social Psychology, 52 (1), 81-90.

Costa, P. T. & McCrae, R. R. (1988). From catalogue to classification: Murray's needs and the Five Factor Model. Journal of Personality and Social Psychology, 52 (1), 81-90.

Costa, P. T. & McCrae, R. R. (1989). The structure of interpersonal traits: Wiggin's Circumplex and the Five-Factor Model. Journal of Personality and Social Psychology, 52 (1), 81-90.

Costa, P.T. & McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychological Bulletin, 104 (2), 163-191.

Cox, R. H. (1988). Utilization of psychomotor screening for USAF pilot candidates: Enhancing predictive validity. Aviation, Space, and Environmental Medicine, 59, 640-645.

Cox, R. H. (1989). Psychomotor screening for USAF pilot candidates: Selecting a valid criterion. Aviation, Space, and Environmental Medicine, 60, 1153-1156.

Damos, D. L. (1993). Using meta-analysis to compare predictive validity of single- and multiple- task measures to flight performance. Human Factors, 35 (4), 615-628.

Damos, D. L. (In press). Pilot selection batteries: Shortcomings and perspectives. International Journal of Aviation Psychology.

Day, D. V. & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. Personnel Psychology, 42, 25-36.

Degani, A. & Wiener, E. L. (1990). Human factors of flight-deck checklists. (NASA Contractor Report 177549). Moffett Field, CA: NASA-Ames Research Center.

Delaney, H. D. (1992). Dichotic listening and psychomotor task performance as predictors of naval primary flight-training criteria. The International Journal of Aviation Psychology, 2 (2), 107-120.

Dessouky, M. I., Moray, N., & Kijowski, B. (1995). Taxonomy of scheduling systems as a basis for the study of strategic behavior. Human Factors, 37 (3), 443-472.

Diehl, A. (1991b). The effectiveness of training programs for preventing aircrew "error". In Sixth International Symposium on Aviation Psychology, Columbus, Ohio, 30 April 1991.

Diehl, A. E. (1991a). Human performance and systems safety considerations in aviation mishaps. The International Journal of Aviation Psychology, 1 (2), 97-106.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. Annual Review of Psychology, 41, 417-440.

Dolgin, D. L., Gibb, G. D., Nontasak, T., & Helm, W. R. (1987). Instructor pilot evaluations of key naval primary flight training criteria. (NAMRL Rep. No. 1331). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Driskell, J. E. & Olmstead, B. (1989). Psychology and the military: Research applications and trends. American Psychologist, 44 (1), 43-54.

Dully, F. E. (no date). The lifestyle keys to flight deck performance of the naval aviator—another window. In C-130 Aircrew Training System Crew Resource Management Workbook.

Endsley, M. R. (1995a). Toward a theory of situational awareness in dynamic systems. Human Factors, 37 (1), 32-64.

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

Endsley, M. R. & Bolstad, C. A. (1994). Individual differences in pilot situational awareness. The International Journal of Aviation Psychology, 4 (3), 241-264.

Fedor, D. B., Rensvold, R. B. & Adams, S. M. (1992). An investigation of factors expected to affect feedback seeking: A longitudinal field study. Personnel Psychology, 45, 779-805.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51 (4), 327-358.

Flynn, C. F., Sipes, W. E., Grosenbach, M. J. & Ellsworth, J. (1994). Top performer survey: Computerized psychological assessment in aircrew. Aviation, Space, and Environmental Medicine, 65 (5, Suppl.), A39-40.

Foushee, C. H. (1982). The role of communications, sociopsychological, and personality factors in the maintenance of crew coordination. Aviation, Space, and Environmental Medicine, 53 (11), 1062-1066.

Foushee, C. H. (1984). Dyads and triads at 35,000 feet. American Psychologist, 39 (8), 885-893.

Fowler, B. (1981). The aircraft landing test: An information processing approach to pilot selection. Human Factors, 23 (2), 129-137.

Frederico, P. (1995). Expert and novice recognition of similar situations. Human Factors, 37 (1), 105-122.

Gibb, G. D. (1990). Initial validation of computer-based secondary selection system for student naval aviators. Military Psychology, 2 (4), 205-219.

Giffin, W. C. & Rockwell, T. H. (1984). Computer-aided testing of pilot response to critical in-flight events. Human Factors, 26 (5), 573-581.

Gnan, M., Flynn, C. F., & King, R. E. (1995). Psychological pilot selection in the U. S. Air Force, the Luftewaffe, and the German aerospace research establishment (Tech . Rep. No. AL/AO-TR-1995-0003).

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. Human Factors, 24, 173-183.

Gopher, D. & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures. Human Factors, 26 (5), 519-532.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. Human Factors, 36 (3), 387-405.

Gregorich, S., Helmreich, R. L., Wilhelm, J. A. & Chidester (1989). Personality based clusters as predictors of aviator attitudes and performance (NASA Technical Memorandum A90-26273). Moffett Field, CA: NASA-Ames Research Center.

Gregorich, S. E., Helmreich, R. L, & Wilhelm, J. A. (1990). The structure of cockpit management attitudes. Journal of Applied Psychology, 75 (6), 682-690.

Hanson, M. A., Hedge, J. W., Logan, K. K., Bruskiewicz, K. T., Borman, W. C., & Siem, F. M. (1996). Application of the critical incident technique to enhance crew resource management training. Eighth International Symposium on Aviation Psychology, Columbus, OH.

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

Hart, S. G. & Hauser, J. R. (1987). Inflight application of three pilot workload measurement techniques. Aviation, Space, and Environmental Medicine, 58, 402-410.

Hartman, B. O. & Secrist, G. E. (1991). Situational awareness is more than exceptional vision. Aviation, Space, and Environmental Medicine, 62, 1084-1089.

Hedge, J. W., Hanson, M. A., Borman, W. C., & Bruskiewicz, K. T. (1994). Examining the feasibility of developing a situational judgment test for Air Force pilots (AL/HR-TR-1994-0065). Brooks AFB, TX: Human Resources Directorate.

Hedge, J. W., Hanson, M. A., Siem, F. M., Bruskiewicz, K. T., Borman, W. C., & Logan,, K. K. (1995). Development and validation of crew resource management selection test for Air Force transport pilots. Eighth International Symposium on Aviation Psychology, Columbus, OH.

Heinrich, Peterson, and Roos (1980). Industrial Accident Prevention. New York: McGraw-Hill.

Helmreich, R. L. (1984). Cockpit management attitudes. Human Factors, 26 (5), 583-589.

Helmreich, R. L. (1986). Pilot selection and performance evaluation: A new look at an old problem. In Proceedings Psychology in the Department of Defense, USAFA-TR-86-1, 1986, April 16-18, USAF Academy, Colorado Springs, Colorado, 274-278.

Helmreich, R. L., Sawin, L. L., & Carsrud, A. L. (1986). The honeymoon effect in job performance: Temporal increases in the predictive power of achievement motivation. Journal of Applied Psychology, 71 (2), 185-188.

Helmreich, R. L. & Wilhelm, J. A. (1991). Outcomes of crew resource management training. The International Journal of Aviation Psychology, 1 (4), 287-300.

Helmreich, R. L., Wilhelm, J. A., Gregorich, S. E., & Chidester, M. A. (1990). Preliminary results from the evaluation of cockpit resource management training: Performance ratings of flight crews. Aviation, Space, and Environmental Medicine, 61, 576-579.

Helton, K. T. & Street, D. R. Jr. (1992). The Five-Factor Personality Model and naval aviation cadets (NAMRL Rep. No. 1379). Pensacola, FL: US Naval Aviation Medical Center.

Hilton, T. F. & Dolgin, D. L. (1991). Pilot selection in the military of the free world published in R. Gal and A. D. Mangelsdorff (Eds), Handbook of Military Psychology. New York: John Wiley & Sons,

Hogan, J. & Broach, D. (1990). Development of a task information taxonomy for human performance systems. Military Psychology, 2 (1), 1-19.

Hughes Training, Inc., Training Division. (no date). C-130 Aircrew Training System Crew Resource Management Workbook (CRI-11801).

Humphrey, D. G. & Kramer, A. F. (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. Human Factors, 36 (1), 3-26.

Hunter, D. R. & Burke, E. F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. The International Journal of Aviation Psychology, 4 (4), 297-313.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96 (1), 72-98.

Hunter, J. E. & Schmidt, F. L. (1992). Dichotomization of continuous variables: The implications for meta-analysis. Journal of Applied Psychology, 75 (3), 334-349.

Hunter, J. E., Schmidt, F. L. & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. Journal of Applied Psychology, 75 (1), 28-42.

Jackson, D. E. & Ree, M. J. (1990). Tool for studying effects of range restriction in correlation coefficient estimation (Tech. Pap. No. AFHRL-TP-90-06). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Jensen, R. S. (1982). Pilot judgment: Training and evaluation. Human Factors, 24 (1), 61-73.

Kanki, B. G. & Foushee, H. C. (1989b). Communication as group process mediator of aircrew performance. Aviation, Space, and Environmental Medicine, 60, 402-410.

Kanki, B. G., Lozito, B. A., & Foushee, H. C. (1989a). Communication indices of crew coordination. Aviation, Space, and Environmental Medicine, 60, 56-60.

King, R. E. (1994). Assessing aviators for personality pathology with the Millon Clinical Multiaxial inventory (MCMI). Aviation, Space, and Environmental Medicine, 65, 227-231.

King, R. E. & Flynn, C. F. (1995). Defining and measuring the "right stuff": neuropsychiatrically enhanced flight screening (N-EFS). Aviation, Space, and Environmental Medicine, 66, 951-956.

Kohen-Raz, R., Kohen-Raz, A., Erel, J., Davidson, B., Caine, Y, & Froom, P. (1994). Aviation, Space, and Environmental Medicine, 65, 323-326.

Koonce, J. M. (1984). A brief history of aviation psychology. Human Factors, 26 (5), 499-508.

Kreienkamp, R. A. & Luessenheide, H. D. (1985). Similarity of personalities of flight instructors and student pilots: Effect on flight training time. Psychological Reports, 57, 465-466.

Layton, C., Smith, P. J., & Mc Coy, C. E. (1994). Design of cooperative problem-solving system for en-route flight planning: An empirical evaluation. Human Factors, 36 (1), 94-119.

Lester, L. F. & Bombaci, D. H. (1984). The relationship between personality and irrational judgment in civil pilots. Human Factors, 26 (5), 565-572.

Li, G. (1994). Pilot-related factors in aircraft crashes: A review of epidemiological studies. Aviation, Space, and Environmental Medicine, 65, 944-952.

Li, G. & Baker, S. P. (1994). Prior crash and violation records of pilots in commuter and air taxi crashes: A case-control study. Aviation, Space, and Environmental Medicine, 65, 979-985.

Linde, C., & Goguen (1991). Checklist interruption and resumption: A linguistic study. (NASA Technical Memorandum 177460). Moffett Field, CA: NASA-Ames Research Center.

Mills, J.G., & Jones, D.R. (1984). The Adaptability Rating for Military Aeronautics: An historical perspective of a continuing problem. Aviation, Space, and Environmental Medicine, 55, 558-562.

Moray, N., Dessouky, M. I., Kijowski, B. A., & Adapathya, R. (1991). Strategic behavior, workload, and performance in task scheduling. Human Factors, 33 (6), 607-629.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. Journal of Applied Psychology, 75 (6), 640-647.

Muckler, F. A. & Seven, S. A. (1992). Selecting performance measures: "Objective" versus "subjective" measurement. Human Factors, 34 (4), 441-455.

Murray, M. W., Siem, F. W., Duke, A. P., & Weeks, J. L. (1995). Dimensions of Air Force Pilot Combat Performance (Tech. Pap. No. AL/HR-TP-1995-0008). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

National Aeronautics and Space Administration. (1996). ASRS celebrates its 20th birthday. Callback, 204, p. 1.

National Transportation Safety Board. (1994). A review of flightcrew-involved major accidents of U.S. air carriers, 1978 through 1990. (Report No. PB94-917001) Washington D.C.: National Transportation Safety Board.

O'Hare, D. (1992). The artful decision maker: A framework model for aeronautical decision making. The International Journal of Aviation Psychology, 2 (3), 175-191.

Olea, M. M. & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. Journal of Applied Psychology, 79 (6), 845-851.

Picano, J. J. (1990). An empirical assessment of stress-coping styles in military pilots. Aviation, Space, and Environmental Medicine, 61, 356-360.

Picano, J. J. (1991). Personality types among experienced military pilots. Aviation, Space, and Environmental Medicine, 62, 517-520.

Platenius, P. H. & Wilde, G. J. S. (1989). Personal characteristics related to accident histories of Canadian pilots. Aviation, Space, and Environmental Medicine, 60, 42-45.

Popplow, J. R. (1994). Selecting, training, and providing operational support to aircrew. Canadian Aeronautics and Space Journal, 40 (2), 51-53.

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

Raymond, M. W. & Royce M. (1995). Aviators at risk. Aviation, Space, and Environmental Medicine, 66, 35-39.

Ree, M. James & Carretta, T. R. (1994). The correlation of general cognitive ability and psychomotor tracking tests. International Journal of Aviation Psychology, 2 (4), 209-216.

Ree, M. J. & Carretta, T. R. (1995). Factor analysis of the ASVAB: Confirming a Vernon-like structure. (Tech. Pap. No. AL/HR-TP-1995-0007). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). A path model of U.S. Air Force pilot training and its antecedents. (Tech. Pap. No. AL/HR-TP-1995-0034). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division

Reinhardt, R. F. (1970). The outstanding jet pilot. American Journal of Psychiatry, 127 (6), 32-36.

Retzlaff, P. D. & Gibertini, M. (1987). Factor structure of the MCMI basic personality scales and common-item artifact. Journal of Personality Assessment, 51 (4), 588-594.

Retzlaff, P. D. & Gibertini, M. (1987). Air Force pilot personality: Hard data on the "Right Stuff", Multivariate Behavioral Research, 22, 383-399.

Retzlaff, P. D. & Gibertini, M. (1988). Objective psychological testing of U.S. Air Force officers in pilot training. Aviation, Space, and Environmental Medicine, 59, 661-663.

Rouse, W. B. & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100 (3), 349-363.

Ruffell Smith, H. P. (1979). A simulator study of the interaction of pilot workload with errors, vigilance, and decisions (NASA Technical Memorandum 78482). Moffett Field, CA: NASA-Ames Research Center.

Salas, E., Prince, C., Baker, D. P. & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. Human Factors, 37 (1), 123-136.

Sanders, M. G. & Hoffman, M. A. (1975). Personality aspects of involvement in pilot-error accidents. Aviation, Space, and Environmental Medicine, 46 (2), 186-190.

Secrist, G. E. & Hartman, B. O. (1993). Situational awareness: The trainability of the near-threshold information acquisition dimension. Aviation, Space, and Environmental Medicine, 64, 885-892.

Selye, H. (1974). Stress without distress. New York , Harper & Row.

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

Siem, F. M. (1992). Predictive validity of an automated personality inventory for air force pilot selection. The International Journal of Aviation Psychology, 2 (4), 261-270.

Siem, F. M. (1995). Optimal personnel assignment: An application to air force pilots. Military Psychology, 7 (4), 253-263.

Siem, F. M. (1996). The use of response latencies to enhance self-report personality measures. Military Psychology, 8 (1), 15-27.

Siem, F. M., Carretta, T. R. & Mercatante, T. A. (1988). Personality, attitudes, and pilot training performance: Preliminary analysis (Tech. Pap. No. AFHRL-TP-87-62). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Siem, F. M. & Murray, M. W. (1994) Personality factors affecting pilot combat performance: A preliminary investigation. Aviation, Space, and Environmental Medicine, 65 (5, Suppl.), A45-48.

Sive, W. J. & Hattingh, J. (1991). The measurement of psychophysiological reactions of pilots to a stressor in a flight simulator. Aviation, Space, and Environmental Medicine, 62, 831-836.

Skinner, J., & Ree, M. J. (1987). Air Force Officer Qualifying Test (AFOQT): Item and factor analysis of Form O (Tech. Rep. No. AFHRL-TR-86-68). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Skogstad, A., Dyregrov, A., & Hellesoy, O. H. (1995). Cockpit-cabin crew interaction: Satisfaction with communication and information exchange. Aviation, Space, and Environmental Medicine, 66, 841-848.

Smith, K. & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. Human Factors, 37 (1), 137-148.

Stoker, P., Hunter, D. R., Kantor, J. E., Quebe, J. C. & Siem, F. M. (1987). Flight screening program effects on attrition at undergraduate pilot training (Tech. Pap. No. AFHRL-TP-86-59). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Stokes, A. F., Banich, M. T., & Elledge, V. C. (1991). Testing the tests - an empirical evaluation of screening tests for the detection of cognitive impairment in aviators. Aviation, Space, and Environmental Medicine, 62, 783-788.

Stokes, A. & Kite, K. (1994). Flight stress: Stress, fatigue, and performance. Cambridge: University Press.

Stone, R. B., Babcock, G. L., & Edmunds, W. W. (1985). Pilot judgment: An operational viewpoint. Aviation, Space, and Environmental Medicine, 56, 149-152.

Street, D. R., Jr. & Dolgin, D. L. (1992a). The efficacy of biographical inventory data in predicting early attrition in naval aviation officer candidate training (NAMRL Rep. No. 1373). Pensacola, FL: US Naval Aviation Medical Center.

Street, D. R., Jr. & Dolgin, D. L. (1993). An evaluation of personality testing and the Five-factor model in the selection of landing craft air cushion vehicle crew members (NAMRL Rep. No. 1385). Pensacola, FL: US Naval Aviation Medical Center.

Street, D. R., Jr. & Dolgin, D. L. (1994). Computer-based psychomotor tests in optimal training track assignment of student naval aviators (NAMRL Rep. No. 1391). Pensacola, FL: US Naval Aviation Medical Center.

Street, D. R., Jr. , Helton, K. T. & Dolgin, D. L. (1992b). The unique contribution of selected personality tests to the prediction of success in naval pilot training (NAMRL Rep. No. 1374). Pensacola, FL: US Naval Aviation Medical Center.

Svensson, E., Angelborg-Thanderz, M., & Sjoberg, L. (1993). Mission challenge, mental workload and performance in military aviation. Aviation, Space, and Environmental Medicine, 64, 985-991.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. Personnel Psychology, 44, 1-26.

Tirre, W. C., Koonce, J. M., & Raouf, K. K. (1994). Ability correlates of skill learning in a flight simulator. Paper presented at the 65th Annual Scientific Meeting, Aerospace Medical Association, San Antonio, Texas.

Tirre, W. C. & Raouf, K. K. (1994). Gender differences in perceptual-motor performance. Aviation, Space, and Environmental Medicine, 65 (5, Suppl), A49-53.

Turnbull, G. J. (1992). A review of military pilot selection. Aviation, Space, and Environmental Medicine, 63, 825-830.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211 (30), 453-458.

Urban, J. M., Bowers, C. A., Monday, S. D. & Morgan, B. B. (1995). Workload, team structure, and communication in team performance. Military Psychology, 7 (2), 123-139.

Verdone, R. D., Sipes, W., & Miles, R. (1993). Current trends in the usage of the Adaptability Rating for Military Aviation (ARMA) among USAF flight surgeons. Aviation, Space, and Environmental Medicine, 64, 1086-1093.

Voge, V. M. (1989). Failing aviator syndrome: A case history. Aviation, Space, and Environmental Medicine, 60(7, Suppl.), A89-91.

Wickens, C. D. & Flach, J. M. (1988). Information processing. In E. L. Weiner and D. C. Nagel (Eds.), Human Factors in Aviation, San Diego: Academic Press, Inc.

Wickens, C. D., Stokes, A., Barnett, B., & Davis, T. (1988). Componential analysis of pilot decision making (Tech. Pap. No. AL-TP-1992-0063). Wright-Patterson AFB, Ohio: University of Illinois Aviation Research Laboratory.

Wiener, E. L. & Nagel, D. C. (1988). Human Factors in Aviation. San Diego: Academic Press, Inc.

Wierwille, W. W. & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. Human Factors, 35 (2), 263-281.

Yacavone, D. W. (1993). Mishap trends and cause factors in naval attrition: A review of naval safety center data, 1986-90. Aviation, Space, and Environmental Medicine, 64, 392-395.

Yerkes, R.M. & Dodson, J.D. (1908). The relationship of strength of stimulus to rapidity of habit formation. Journal of Comparative Neurology, 18, 459-482.

Appendix A

NEO Personality Inventory - Revised

1. *NEUROTICISM*

   Anxiety

   Angry Hostility

   Depression

   Self-consciousness

   Impulsiveness

   Vulnerability

2. *EXTRAVERSION*

   Warmth

   Gregariousness

   Assertiveness

   Activity

   Excitement-Seeking

   Positive Emotions

3. *OPENNESS*

   Fantasy

   Aesthetics

   Feelings

   Actions

   Ideas

   Values

4. *AGREEABLENESS*

Trust

Straightforwardness

Altruism

Compliance

Modesty

Tender-Mindedness

5. *CONSCIENTIOUSNESS*

Competence

Order

Dutifulness

Achievement Striving

Self-Discipline

Deliberation

Revised NEO Personality Inventory (NEO-PI-R; Costa and McCrae, 1992)

Appendix B

Peer Survey Rating Categories

1. GENERAL KNOWLEDGE
    - possesses a good fund of information
    - absorbs new information quickly
    - reduces complex issues to essential elements
    - valued for opinions on technical matters
2. JOB PERFORMANCE
    - accomplishes any task thoroughly and efficiently
    - uses initiative to solve difficult problems
    - is predictable, consistent, and reliable in performance
    - able to prioritize multiple critical tasks quickly
3. STRESS TOLERANCE
    - demonstrates prompt and accurate reactions
    - effective in an unexpected emergency; effective under prolonged periods of stress
    - arrives at practical conclusions in emergencies
4. LEADERSHIP
    - motivates others to complete tasks
    - delegates work and allows person to complete task
    - is decisive/flexible when required
    - has determination and projects decisiveness
5. GROUP COHESIVENESS
    - puts group goals ahead of individual goals
    - shares credit and accepts blame
    - tolerant of individual/cultural differences
    - works effectively with many different people
6. TEAMWORK
    - easy to get along with, good sense of humor
    - pulls own weight and does own share of undesirable tasks
    - gives and accepts feedback/criticism well
    - good listener
7. PERSONALITY
    - tolerates difficulties and frustration well
    - few irritating qualities; personable and amiable
    - self-sufficient, motivated, self-starter
8. COMMUNICATION SKILLS
    - presents self well and speaks clearly and effectively
    - represents squadron well; concise and focused
    - gets point across
9. AGGRESSIVENESS
    - pursues goals, rather than waiting for them to occur
    - accepts calculated risks
    - makes opportunities where few seem to exist
    - desire to excel

Rating Categories (Flynn, Sipes, Grosenbach, & Ellsworth, 1994)