Age 60 Study, Part II: Airline Pilot Age and Performance – A Review of the Scientific Literature

AD-A286 246

D.T. Hyland
E.J. Kay
Lehigh University
Bethlehem, Pennsylvania

J.D. Deimler
Hilton Systems, Inc.
Cherry Hill, New Jersey

E.B. Gurman
University of Southern Mississippi
Hattiesburg, Mississippi

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

DTIC
ELECTE
NOV 0 9 1994
D

October 1994

Final Report

94-34433

DTIC QUALITY INSPECTED 5

U.S. Department
of Transportation

**Federal Aviation
Administration**

94 11 4 081

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof.

1. Report No. DOT/FAA/AM-94/21	2. Government Accession No	3. Recipient's Catalog No
4. Title and Subtitle Age 60 Rule Research, Part II: Airline Pilot Age and Performance--A Review of the Scientific Literature		5. Report Date October 1994
		6. Performing Organization Code
7. Author(s) D. T. Hyland, Ph.D., E. J. Kay, Ph.D., J. D. Deimler, E. B. Gurman, Ph.D.		8. Performing Organization Report No 8025-1A
9. Performing Organization Name and Address Hilton Systems, Inc. 950 North Kings Highway Suite 202 Cherry Hill, New Jersey 08034		10. Work Unit No (TRAS)
		11. Contract or Grant No DTFA02-90-C-90125
12. Sponsoring Agency name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code

15. Supplementary Notes

This document is one of four products completed as a part of the Age 60 Rule research contract monitored by Pam Della Rocco, Civil Aeromedical Institute, Contracting Officer's Technical Representative. This work was performed under Task AM-A-91-HRR-138.

16. Abstract

This review of the literature establishes the scientific foundation for subsequent studies on the Age 60 Rule research conducted under a contract with Hilton Systems, Inc. The scientific literature relevant to the two separate scientific approaches required by the contract is reviewed. The document first provides a review of the "Age 60 Rule," as well as the theoretical and methodological considerations critical to the study of aging and the assessment of individual pilot performance. A proposed model is presented to form a framework for the research. Pilot behaviors affected by aging and/or experience are reviewed. Specific performance assessment batteries are reviewed in detail in Appendix A. Issues related to measuring complex pilot performance are discussed. Recommendations and criteria for developing a performance methodology are presented. The second part of the literature review provides a discussion of the issues related to the analyses of existing data to assess the relationship between pilot age, experience, and accident rates. This section provides a critical review of existing analyses and presents recommendations for an improved analytical methodology.

17. Key Words Pilot Age and Performance Age 60 Rule Mandatory Pilot Retirement	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 136	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

PREFACE

This technical report, entitled "Age 60 Rule Research, Part II: Airline Pilot Age and Performance--A Review of the Scientific Literature," is the second in a collection of four major documents resulting from a two year contract to scientifically examine issues related to the Federal Aviation Administration's (FAA's) mandatory retirement regulations for pilots. The first report, entitled "Age 60 Rule Research, Part I: Bibliographic Database," was published as an Office of Aviation Medicine Technical Report (DOT/FAA/AM-94/20).

The Federal Aviation Regulations (FARs), Part 121, prohibit individuals from serving as captain or copilot (1st officer) of an aircraft in air carrier operations if those persons have reached their 60th birthday. Commonly referred to as the "Age 60 Rule", the regulation was implemented in response to concerns about the safety of aging pilots as the airline industry transitioned into the jet age. Although the rule has withstood legal and legislative challenges, little scientific evidence has been available to either support the rule or to guide the FAA to an appropriate alternative.

In 1990, the FAA's Associate Administrator for Certification and Regulation (AVR-1), Mr. Anthony Broderick, requested and sponsored a two year research contract to examine the relationship between age, experience, and accident rates. The Civil Aeromedical Institute (CAMI) was assigned the task of developing and monitoring the contract. In September 1990, the contract was awarded to Hilton Systems Inc., of Cherry Hill, New Jersey. Hilton Systems collaborated with Lehigh University faculty to supplement technical expertise. The FAA requested that the contractor engage in a fresh, innovative approach to issues involved in the Age 60 Rule.

Two separate scientific approaches to the research were requested in the contract's Statement of Work as follows: 1) assessment of the group-based standard (Age 60 Rule) through investigation of the relationship between age, experience, and accident rates from existing data bases, and 2) examination of the feasibility of developing an individually-based performance assessment battery. This literature review served to establish the scientific basis for both investigative approaches in the contract.

Finally, for clarity of presentation, graphs presented in Appendix B are Hilton Systems' reproductions of Golaszewski's figures from his original data. Some computational errors were corrected in the Hilton reproductions. Differences in the figures provided here and Golaszewski's report are due to these modifications. In addition, the x-axis in several of the Hilton Systems' graphs were labeled differently.

Pamela Della Rocco, COTR

CONTENTS

APPENDICES

  A. Test Batteries Used for Pilot Selection or Evaluation

  B. Figures Plotting Age, Flight Time, and Accidents from
     Golaszewski (1983)

  C. Topical Bibliography of Aging, Medical Variables, and
     Pilot Performance

Accesion For		
NTIS CRA&I	☑	
DTIC TAB	☐	
U:.announced	☐	
Justification		
By		
Di∴t ib∴tion /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

1. INTRODUCTION

The "Age 60 Rule", contained in Part 121 of the Federal Aviation Regulations (121.383C), mandates retirement at age 60 of commercial airline pilots-in-command and co-pilots. When the rule was established in 1959, the stated aim was to reduce pilot contribution to aircraft accidents. This emphasis on pilot fitness and performance was assumed to be increasingly critical because of the industry's shift to aircraft with increased speed and passenger-carrying capacity. The selection of age 60 as a mandatory retirement standard was based on available studies which indicated that with increasing age there was progressive deterioration of relevant physiological and psychological functions. These studies also indicated that increasing age was associated with increased incidence of diseases such as heart attack and stroke, which were considered particularly problematic because of their association with sudden incapacitation. While there was recognition of the fact that not all individuals experience equivalent age-related deteriorations in health and performance, an age limitation was deemed necessary because deterioration in performance could not be reliably and objectively measured in individuals.

In 1979, the US Congress mandated that the Age 60 Rule be re-evaluated. The general aim of this evaluation was to assess the effect of aging on the ability of individuals to perform the duties of pilots with the highest level of safety (National Institute on Aging, 1981). The National Institute on Aging (NIA) based their evaluation on an Institute of Medicine (IOM) committee assessment (1981) of existing relevant scientific data and comments on the IOM report. The NIA panel was not able to identify a medical or performance appraisal system that could identify those pilots who would pose a safety hazard because of early or impending deterioration in health or performance, and therefore they recommended that the Age 60 Rule be retained. However, they also recommended that systematic collection of medical and performance data needed to further evaluate the Age 60 Rule be carried out. With respect to medical data they recommended more stringent and under certain conditions more frequent screening of physical health and in particular cardiovascular health in light of the increasing risk of cardiovascular accidents with age. With respect to performance data, they pointed out that it is possible to test aspects of pilot performance in simulators which can reproduce critical flight situations. However, they were aware of no available standard for grading performance in objective quantifiable ways. Therefore, the NIA panel recommended that:

- Additional research be conducted to develop quantifiable, objective criteria for measuring overall pilot performance; and

- Performance data be systematically collected to further evaluate the Age 60 Rule.

Ten years later, despite increased research attention being paid to these issues, these two recommendations have not been sufficiently met. Thus, these recommendations are the basis of the present Age 60 Project.

The ultimate long-term aim of the Age 60 Project is to enhance aviation safety by increasing understanding about the relationships among pilot age, experience, performance, and accident rates. The specific aims of the present project are:

- To consolidate the currently separate data bases related to pilot certification and aviation accidents and incidents, and to use this consolidated database to statistically examine the relationships that exist among chronological age, recent flight time, total flight time, and accidents (Task 3);

- To develop a methodology for the assessment of individual pilot performance (Task 4);

- To plan a longitudinal study that would systematically examine the effect of aging on pilot performance (Task 5).

The empirical evidence obtained in this project concerning the relationships among pilot age, experience, and accident rates and the effects of age and experience on pilot performance should be extremely relevant to an evaluation of the Age 60 Rule. Evaluation of these relationships is an important step in determining the feasibility, reliability, and validity of individual pilot performance assessment as an alternative to the Age 60 Rule. However, a definitive stance on the Age 60 Rule will also require consideration of pertinent issues beyond the scope of the present tasks. For example, before the Age 60 Rule can be replaced with an individual retirement standard, the medical certification procedure may need to be enhanced to more stringently certify that the pilot is free from impairing physical and mental disease. This study will also not explicitly address in detail issues related to the political, economic, pension planning, pilot advancement, or age discrimination issues that have entered into the Age 60 Rule debate.

The present report reviews the scientific literature relevant to the three major tasks of this project. The organization of this report is as follows. Section 2 will examine the influence of aging on pilot performance. First, a conceptual approach for assessing individual pilot performance and for examining the influence of aging on this performance will be presented. Data related to the components of the model will be reviewed. Finally, conclusions and recommendations for a pilot performance assessment methodology (Task 4 of the present project) and for the longitudinal examination of the effects of aging on pilot performance (Task 5) will be presented. Section 3 will consider issues bearing on the relationships among pilot age, recent and total flight time, and aviation accidents (Task 3 of the present project). Emphasis will be placed on a critical study by Golaszewski (1983), paying particular attention to the methodology and how to improve it. Since, the consolidated database will combine information about accidents with information about the medical status of the pilots, a brief review of the literature pertinent to the relationships among accidents, medical condition, and age will be presented, and ways in which the consolidated database could shed light on these issues will be suggested.

2. AGING AND THE ASSESSMENT OF INDIVIDUAL PILOT PERFORMANCE

2.1 Specific Aims

This section of the report will review scientific literature relevant to Tasks 4 and 5 of the Age 60 Project. Task 4 is aimed at development of a methodology to quantitatively and

objectively assess an individual pilot's performance. Specifically, these three levels of performance will be considered:

- Basic psychomotor, perceptual, and cognitive skills which are not specifically related to flying but are assumed to influence pilot performance;

- Knowledge, judgment, and decision making related to piloting an aircraft;

- Pilot performance in a simulator under varying flight conditions.

Task 4 will examine relationships among these performance measures within two age groups. The data obtained should be useful in the following respects:

- Provide a preliminary step in the construction of an individual performance measure that could eventually replace the Age 60 Rule.

- Development of an assessment methodology for use in a longitudinal examination of the effects of aging on pilot performance.

- Preliminary cross-sectional examination of age differences on these measures of pilot performance.

Task 5 is aimed at planning a longitudinal examination of the effects of aging on pilot performance. Key questions to be addressed in this study are:

- With increasing age are there declines in performance on measuresof basic skills, pilotknowledge and decision making, or simulated flying performance? At what age do the declines occur? How much individual variation characterizes these age changes?

- Are there increments in performance on any of the above measures associated either with pilots' increasing age or experience?

- Are declines in some aspects of pilot performance being compensated for by increments in other aspects of pilot performance? To what degree? Under what conditions? What is the mechanism? For example, if flying performance in the simulator appears unaffected by increasing age, is it because age declines in basic skills are compensated for by increases in pilot knowledge?

To accomplish the research objectives of Tasks 4 and 5 of the Age 60 Project, two preliminary steps must be taken. Step 1 is to develop a high-level conceptual model for assessing the relationships among aging, experience, and pilot performance. This model will be described below in section 2.2. Step 2 is to review the scientific literature in the areas of cognitive aging and pilot performance within the framework of this conceptual model. This review is presented in section 2.3.

3

2.2 Conceptual Approach

2.2.1 Existing Models of Aging and Performance

Advancing age is associated with declines in physiological functioning and an increase in the prevalence of disorders. However, individuals vary considerably in their rates of physiological aging and in their predisposition to disease. Thus, while average levels of functioning may decrease with increasing age, individual variability appears to increase. Behavioral researchers studying age have found it useful to draw a conceptual distinction (although in practice this may be difficult to diagnose) between: primary aging, which refers to normal disease-free aging; and secondary aging, which refers to disease-related aging. While primary aging reflects changes inherent to the aging process itself that are ultimately irreversible, secondary aging reflects changes that are not inevitable or always irreversible. Rather than being a result of the aging process itself, secondary aging changes may reflect a genetic predisposition or they may be lifestyle related.

In addition to distinguishing primary and secondary aging, researchers have also found it useful to distinguish usual patterns of aging (that describe the average or typical older person) from successful patterns of aging that describe the person who shows lesser declines in functioning. For example, factors such as internal locus of control, social support to deal with life stressors, a healthful diet, and exercise (Rowe & Kahn, 1987) seem to be associated with more "successful" aging. An active, athletic 60 year old may "seem younger" than a sedentary 40 year old smoker. This is the essential problem with using chronological age as a meaningful marker. Chronological age implicitly involves a mixing of these influences. It reflects both primary and secondary aging and does not distinguish individuals who display typical as opposed to successful patterns of aging.

Despite these problems in using chronological age as a meaningful marker of an individual's functioning, it is still important to investigate the relationship between increasing age and functioning, at least in terms of group patterns. Researchers have attempted to identify typical patterns of change associated with increasing age for a variety of aspects of individual functioning. Schaie (1973; 1977; 1986) describes four alternative trend models of changes with aging:

- The **increment model** refers to behaviors that increase with age. It is used extensively to describe child development. There may be some aspects of pilot performance that also match this increment model (i.e., knowledge related to piloting an aircraft).

- The **stability model** predicts no change with age in disease-free individuals. Many aspects of personality may follow a stability model. For example, for pilots some aspects of personal style related to crew coordination and leadership may not change in systematic ways with age.

- The **irreversible decrement model** applies to processes where there is direct biological control over age-related decline in performance. These changes are

4

assumed to be unaffected by environmental factors such as experience. Psychomotor slowing is a good example.

- In the **decrement with compensation model**, environmental inputs may partially compensate for biologically determined decrements to produce stability or even increments in abilities. For example, there are anecdotal reports of some older pilots performing exceedingly well under extreme emergency situations. Perhaps, their experience and increased knowledge compensated for any age-related psychomotor slowing that they may have experienced.

No one model is likely to explain all aspects of a complex behavior like piloting. However, since one goal is to determine if there are age-related changes that are associated with deterioration in piloting skills, there must be a search for aspects of performance that fit a decrement model. Second, it is important to consider the concept of compensation. Is compensation possible in a complex behavior such as pilot performance? To what degree? Under what circumstances? What is the compensatory mechanism?

Some researchers (Charness, 1983; Rybash, Hoyer, & Roodin, 1987; Salthouse, 1987) studying age differences in complex behaviors such as typing, chess, bridge, physics, and computer programming have found that older people seem to do as well as young people on these complex tasks. Surprisingly, these older adults seem to have experienced average age-related declines on the component psychological skills that are related to the complex behavior on which they are successful. In other words, there appears to be compensation for declines in the basic skills. Of course this compensation is only possible when the older adult has developed expertise in the complex task. An older novice would not be able to make use of compensation, and in fact would tend to perform more poorly than a younger novice because of the age-related declines in basic skills. Thus, expertise (knowledge in a particular domain that has become intuitive, automatic and highly skilled) plays a crucial role in the compensation process.

Rybash, Hoyer, and Roodin (1987) propose a theory of what they term encapsulation-the process of compensating through expertise for age-related declines in intellectual or cognitive functioning. Encapsulation is the theoretical notion that the domain-independent processes of cognition such as attention, memory and reasoning become linked with the product of cognition-knowledge within a specific domain. As a result of this linkage, performance in a domain of expertise may be unrelated to declines in domain-independent cognitive processes because of the compensating role of expertise (the domain-dependent knowledge base). For example, it is possible that domain-dependent knowledge and judgment related to flying could compensate for age-related declines in the older pilot's domain-independent processes such as attention, information processing speed, and working memory. Thus, to understand aging and pilot performance it is necessary to look not only at cognitive processes but also the domain in which they are encapsulated.

Finally, it is important to determine in what domains, and to what degree, expertise can compensate for decrements in domain-independent processes. For example, Charness (1985) and Rybash, Hoyer, and Roodin (1987) suggest that the cumulative effects of age and experience that enhance the expert's knowledge base are most likely to compensate in certain domains, such
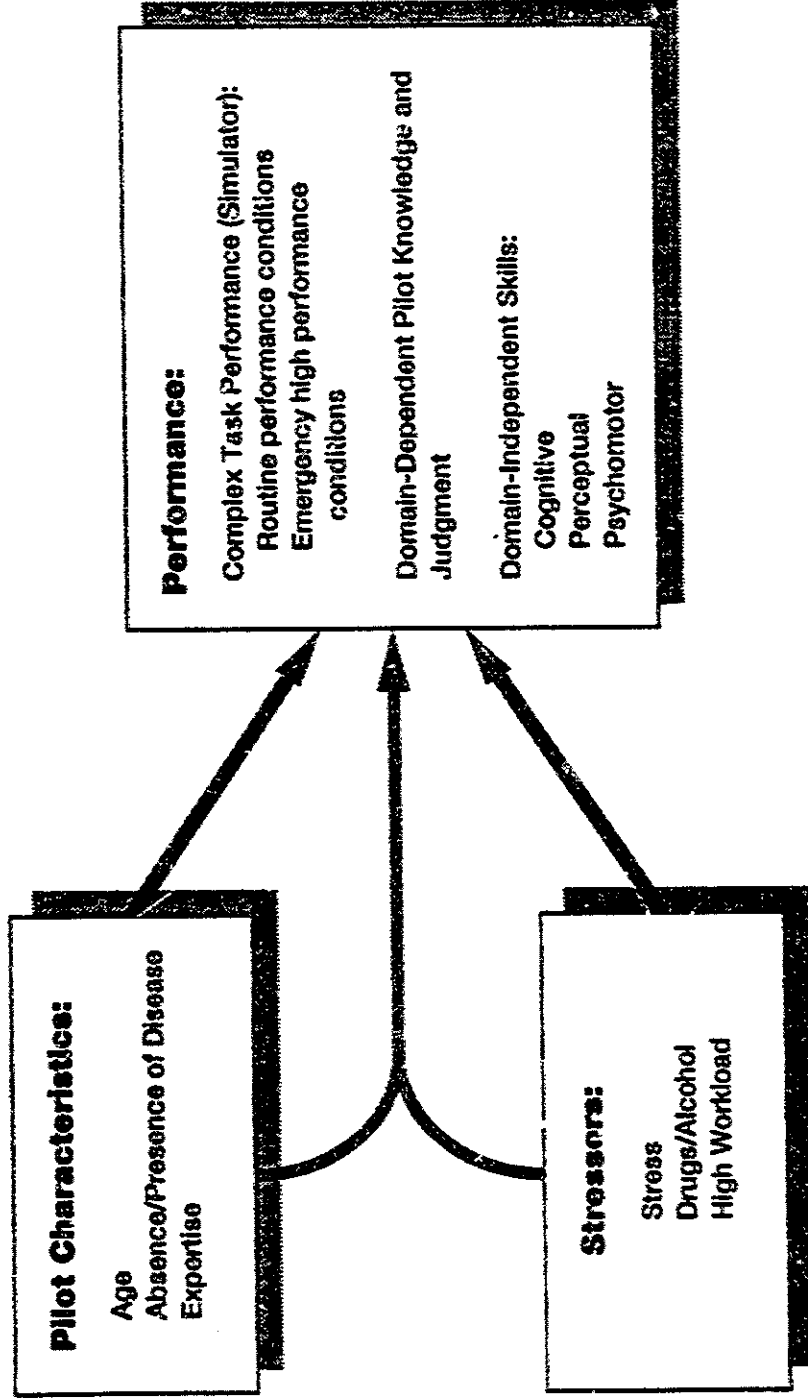
5

as chess or computer programming, where performance allows more time for predictability, planning ahead and reflection, and demands fewer snap decisions and less physical exertion. In the latter domains, such as performance in sporting activities like basketball, expertise cannot totally compensate. Performance declines with age even among those individuals who display expert performance during young adulthood. Thus, the degree to which older pilots can compensate for declines in basic skills likely depends on the type of flying they are doing. Under normal, routine conditions flying may be more similar to "chess" and compensation may be possible. Under extreme emergency conditions the performance demands (e.g., quick decision-making and response) may become more similar to "basketball" and compensation may no longer be sufficient to maintain the same level of complex task performance.

2.2.2 A Proposed Model for the Assessment of Aging and Pilot Performance

While there is a wealth of data on the effects of aging on perceptual, psychomotor, and cognitive abilities, most of this research involves basic skills tested under laboratory conditions. Comparatively little attention has been directed at the effects of aging on complex behavior in real world situations, such as flying an aircraft. There is a large and burgeoning literature on human factors aspects of piloting an aircraft, yet comparatively little aimed directly at the effect of aging on pilot performance. Because of the relatively small number of studies that have directly investigated the effects of aging on specific aspects of pilot performance, there is a great need for systematic investigation of multiple aspects of pilot performance across a wide age span. In order for this investigation to be most fruitful, it should be guided by a coherent conceptual framework. The framework proposed for the present project (see Figure 1) attempts to draw together the fragmented research on aging and pilot performance from the cognitive aging literature and aviation human factors literature into a model of aging and pilot performance that is similar to the decrement with compensation models that have been developed in the area of cognitive aging and that have successfully contributed to understanding the concept of expertise. The model is composed of three classes of variables:

- **Pilot characteristics**, including age, health status, and amount and type of flying experience. It is important to consider health status as well as age to attempt to disentangle as much as possible the effects of primary and secondary aging. Pilots' medical certification procedures screen out cases of serious health impairment, but does more subtle, subclinical disease affect performance? It is also important to consider the role of experience. To what degree does experience itself (i.e., simple practice as measured in flight hours) correlate with levels of expertise? Charness (1989) points out that experience in the form of practice is necessary though not sufficient for developing high levels of expertise and skill. All Part 121 pilots have a minimal level of experience which is significant, but does increasing this level of experience have beneficial effects? Up to what point?

- **Stressors**, such as alcohol or drugs, circadian disturbances, general life stress, and increased workload. These stressors may influence pilot performance directly and/or interact with pilot characteristics to influence performance. For example, alcohol and other drugs may have a greater affect on pilot performance in older,

6

FIGURE 1. A Model to Assess Age-Related Changes in Pilot Performance

Pilot Characteristics:

Age
Absence/Presence of Disease
Expertise

Stressors:

Stress
Drugs/Alcohol
High Workload

Performance:

Complex Task Performance (Simulator):
Routine performance conditions
Emergency high performance
conditions

Domain-Dependent Pilot Knowledge and
Judgment

Domain-Independent Skills:
Cognitive
Perceptual
Psychomotor

compared to younger, pilots (Collins & Mertens, 1988: Leirer, Yesavage, & Morrow, 1989; Morrow, Leirer, & Yesavage, 1990).

- **Pilot performance**, which includes basic cognitive, perceptual and psychomotor skills; flight-relevant knowledge and judgment; and flight performance in a simulator under varying flight conditions. Changes in domain-independent basic skills may not show a relationship to complex performance until a critical threshold is exceeded. These thresholds may themselves change with age. One way to explore this situation is to identify younger and older subjects who are matched on the complex behavior (correlation between performance and age is zero); then break down the components of the complex behavior to show age-related decline in some component that is compensated for by enhanced skill in another component. For example, declines in basic skills could be compensated for by increases in pilot knowledge and judgment that occur as pilots' expertise increases. However, increased knowledge may only compensate for declining skills under ideal or routine task conditions. Therefore, pilots' performance must be tested under a range of flying conditions. Standard maneuvers, because they are so well-learned and practiced, may only be useful in detecting obvious decrements in pilot performance. Most of the literature on age and performance suggests that as tasks become more complex and demanding, age-related performance declines are more pronounced. Cerrella, Poon, and Williams (1980) call this robust finding the age-complexity hypothesis. Performance of complex maneuvers under stress or in novel situations is more likely to be impaired in older pilots than well-learned familiar tasks (National Institute on Aging, 1981). Emergency flight conditions such as turbulence, bad weather, and increasing communication demands could be manipulated and pilots' coping assessed.

2.3   Background Literature

2.3.1 Pilot Characteristics: The Aging Pilot

Within the separate research areas of cognitive aging and pilot performance, there exists a vast literature relevant to the effects of aging on pilot performance. However, the intersection of these two areas of research is composed of a very small number of studies that directly examine the effects of aging on pilot performance. Therefore, while the present review will describe in detail the methods and results of studies that have examined aging pilots, it will also be necessary to make reference to studies within the area of cognitive aging that have examined non-pilot subjects performing more generic laboratory tasks. Prior to reviewing this research, it is important to note several general issues that must be considered in interpreting these findings with respect to the current project.

First, the few studies done on aging pilots are generally limited in three respects: the pilots studied, the research design, and the aspects of performance examined. The research that has been done with aging pilots frequently utilized private pilots as subjects. Those results may not generalize well to professional pilots with thousands of hours of flight experience. In addition, the scarce data on the effects of aging on pilot performance is predominantly

cross-sectional. That is, the studies have compared (at a single point in time) groups of pilots that vary in age, rather than repeatedly testing the same group of pilots as they age (a longitudinal design). While there are interpretation difficulties with both of these simple research designs, a reliance on cross-sectional data is particularly difficult to interpret because different age cohorts of pilots can vary on a number of dimensions besides age (such as health, education, and experience). Finally, the few longitudinal studies that have been conducted on pilots, including The One-Thousand-Aviator-Study (MacIntyre, Mitchell, Oberman, Harlan, Graybiel, & Johnson, 1978) and The Lovelace Foundation Study (Proper, 1969), did not emphasize cognitive abilities nor correlate these abilities with in-flight performance.

Extreme care must also be taken in generalizing the findings with respect to cognitive aging in non-pilots performing generic laboratory-based cognitive tasks to the flying proficiency of the aging aviator. Generalization is frequently limited due to issues related to subject selection and task selection. Pilots may represent a select population that is systematically different than the general population of older adults. For example, older subjects who are drawn from the general population of community-residing elderly may on average be less educated and less physically fit than older pilots due to pilots' initial selection procedure and continuing medical certifications (Institute of Medicine, 1981). Many of these studies are also based on comparisons between extreme age groups who may not be equated on other key variables (i.e. comparison of 20 year-old college students and 70 year-olds attending a senior citizens activity center). Literature in the area of cognitive aging has also emphasized laboratory tasks. A reliable age difference in a laboratory reaction time task that involves a few hundred milliseconds may not have any significance in the aircraft (Institute of Medicine, 1981). Unfortunately, the type of cognitive aging data that would be most generalizable to aging and pilot performance are not available. There are very few studies that have examined the effects of aging on the performance of complex professional skills and even fewer studies that examine the contribution of practice or experience to the performance of these complex skills.

It would be beyond the scope of the present task to comprehensively review the existing literature on cognitive aging and comprehensively review the aviation human factors literature. Therefore, the scope of the present review will be limited to a brief review of the few studies that have comprehensively examined the effect of aging on pilot performance and then a more systematic evaluation of the discrete cognitive functions related to pilot performance that may be affected by aging.

The few studies that have directly and comprehensively examined the effects of aging on pilot performance are of two distinct types. The first category of studies (Bohannon, 1969; Gerathewohl, 1977; 1978a; Institute of Medicine, 1981; Tsang, 1990) is based on critical review, analysis, and integration of existing research, rather than actual data collection. When appropriate, the conclusions of these review articles are incorporated into the present literature review.

The second category of studies of the effects of aging on pilot performance (Braune & Wickens, 1984a; 1984b; Szafran, 1966; 1969) is aimed at the development of a "functional age profile" that would address the issue of interindividual and intraindividual differences in rate of aging. "Certain individuals may 'age' more rapidly in some or all components of performance than others. Hence, a given individual may be functionally old or young, independent of his or

9

her chronological age" (Braune & Wickens, 1984a, p.3). Therefore, these data collection studies assess a range of psychological abilities presumed to be important to pilot performance and correlate such data with chronological age and pilot performance.

Szafran (1966; 1969) investigated cross-sectional age differences (airline, military and test pilots were selected to represent each decade from age 20 to 60) in perceptual and psycho-physiological skills. Each pilot was administered a battery of tests that was designed to measures the following flight-related skills: making high-speed decisions, detecting low probability and low intensity signals, and retaining significant amounts of information (Szafran, 1966). For example, pilots completed sensory perception tasks measuring visual accommodation, Critical Flicker Fusion (CFF), dark adaptation, and hearing thresholds. They also completed a serial-choice reaction time task by itself (low workload) and simultaneously with a short-term memory task (high workload). The conclusion for almost all measures studied was that pilots' increasing age was not related to performance of these flight-related tasks. Age-related performance decrements were only found in the high workload condition which required the simultaneous performance of two tasks.

Braune and Wickens (1984a) studied 100 non-pilot subjects between age 20 and 60 and had them complete a battery of information processing, aviation-relevant tests. They identified three general age-related changes:

(1)   Some perceptual-motor and spatial tasks (e.g., second-order tracking, maze tracing, dichotic listening) showed monotonic decline across the entire age span.

(2)   A second set of tasks (e.g., visual-spatial and visual-verbal Sternberg memory search tasks), primarily reflecting processing speed, declined systematically, but only after age 40.

(3)   A third set of tasks, primarily reflecting time-sharing, showed no influence of aging. Decline with age in dual-task performance was not greater than decline with age in single task performance.

They also found that there was large variability within groups, with some older subjects displaying performance superior to the mean of the young group. In a follow-up study (1984b), Braune and Wickens studied a sample of 30 male pilots between the ages of 20 and 60. They gave the subjects a similar battery of information processing aviation-related tests and then had each perform maneuvers in a GAT-2 twin engine simulator while simultaneously completing a communications task. They found that 24% of the variance on the flight simulator task and 46% on the communications task could be captured by the age-sensitive information-processing tests.

The remainder of this section will examine the cognitive functions considered to be crucial to piloting. The selection of these functions was based on existing conceptualizations of the abilities underlying pilot performance (Braune & Wickens, 1984a; Gerathewohl, 1977; 1978a; 1978b; Imhoff & Levine, 1981; North & Gopher, 1976). The constructs and processes listed in Table 1 were found in one or more of these taxonomies. Review of pilot and non-pilot studies related to each of these constructs follows. For each construct, relevant data that addresses the question of the influence of aging and/or experience on that construct will be

TABLE 1. Key Piloting Behaviors Affected by Aging and/or Experience

Construct	Process	Sample Piloting Activity
Perceptual	Auditory Perception	Listen to ATC communication, listen for variance in engine noise or air flow over flight control surfaces
	Visual Perception	Inspect, read instruments, observe changes in instrument readings, detect and follow runway lighting
	Visual Scanning	Instrument scan, air traffic detection
Psychomotor	Tracking	Track instrument and external sources and adjust controls appropriately
Attention	Selective/Focused	Ignore task-irrelevant information
	Divided	Multiple task processing (e.g., reading instruments during ATC communication sequence)
	Switching	Attend to several subsystem states during emergency to assess situation
	Sustained	Maintain constant monitoring of instruments and external cues (especially for situation awareness during long flights)
Memory	Short term and working memory	Maintain information for immediate processing (e.g., runway clearance information)
	Long term memory	Remember standard operating and emergency procedures
	Prospective	Remember to perform standard scans and safety checks
Information Processing	Speed	Identify inputs quickly (e.g., recognize anomalous condition which could lead to an emergency)
	Visual spatial	Visualize 3-D situation, modification or rotation of mental image for awareness of aircraft location and direction, distance, and speed perception
	Verbal sequential	Process ATC communication (e.g., for awareness of surrounding air traffic)
Problem Solving and Decision-Making	Domain Independent	Ability to apply rules, ability to draw conclusions regarding interrelationships among variables, mental arithmetic.
	Domain Dependent	Assess situation and decide on appropriate action (e.g., interpret how weather impacts flight plan); identify and evaluate the worth of alternative solutions (e.g., mechanical failures)

11

reviewed. While for ease of presentation each function is discussed separately, it should be noted that in reality these functions are closely interrelated and jointly influence various aspects of pilot performance. For example, pilot difficulty in decision making may reflect both a slowing in processing time and and a memory deficit (Tsang, 1990).

Perceptual abilities. Banich, Stokes and Elledge (1988) reviewed relevant literature on aging and performance of visuo-spatial tasks (such as facial discrimination, and judgments of line orientation) and concluded that advancing age is associated with performance declines. This age-related decline has also been found in pilots. The older pilots in Braune and Wickens (1984b) study performed more slowly on a hidden figures task than the younger pilots did. (Pilots studied were 20 to 60 years old).

The literature not only indicates that there are age differences for some types of perceptual behavior, but also has demonstrated links between performance on these types of perceptual tasks and performance of complex behaviors (automobile driving) that may be relevant to pilot performance. For example, field dependence, an aspect of perceptual style that reflects an individual's susceptibility to the effects of interfering stimuli, has been found to be greater in older than younger individuals and has been found to be associated with impaired automobile driving performance (Institute of Medicine, 1981). Temple (1989) also found that older adults' (aged 60-78) driving performance was significantly related to their performance on perceptual tasks. Aging has also been found to be associated with poorer performance in estimating vehicle velocity (Scialfa, Guzy, Leibowitz, Garvey, & Tyrrell, 1991), and slower reaction to symbolic traffic signs (Halpern, 1984). In both these studies, the older subjects ranged from 60 to 75 years of age and performed more poorly in comparison to subjects under 30 years of age.

Psychomotor skills. "Sensory, perceptual, decision, and motor processes are involved in psychomotor performance. All are more or less affected by aging" (Gerathewohl, 1977, p. 5). Braune and Wickens (1985), who studied subjects ranging in age from 20 to 60, found that second-order tracking performance, and latency of maze tracing significantly declined with age. Mertens, Higgins and McKenzie (1983) also found age decrements (in 60-69 year-olds compared to 40-49 year-olds and 20-29 year-olds) on a psychomotor tracking task. These age-related decrements were more pronounced at moderate and high workload levels. These age-related declines in performance on psychomotor tasks such as tracking tasks may to a significant degree reflect age-related differences in speed of processing. Longitudinal studies show a similar pattern of age-related slowing as the cross-sectional studies (Fozard, Vercruyssen, Reynolds, & Hancock, 1990).

This is an important area to investigate not only because of the apparent age declines but also because psychomotor skills have been found to correlate with the performance of complex behaviors such as driving and to be predictive of automobile accidents (Barrett et al., 1977). In addition, examination of aviation accident statistics indicates that faulty perceptual-motor behavior was a causal factor in a large proportion of nonfatal pilot-caused accidents (Gerathewohl, 1978b).

However, it is also important to note that there are large individual differences in psychomotor slowing and that this type of performance is highly influenced by practice (Institute

of Medicine, 1981). Braune and Wickens (1984b), studying instrument-rated pilots aged 20 to 60, found that tracking performance was significantly correlated with experience but not with age. Thus, practice and experience may counteract the detrimental effects of aging. A study by Stones and Kozma (1989) provides strong evidence that the overpracticing of an activity can counteract the effects of aging on psychomotor performance. They compared young and old subjects in terms of the number of times they could tap their hands and their feet in 30 seconds. Large age differences were found for hand tapping but not foot tapping. Hand tapping is an unpracticed task; foot tapping incorporates movements practiced everyday during normal locomotion.

Attention. Braune and Wickens (1984a) point out that the pilot must continually monitor visual and auditory information from the external environment and within the cockpit. Because of this need to monitor multiple sources of information, "the pilot must be able to focus on the critical cues (selective attention), switch attention as different cues become relevant at different times (switching attention), and often integrate cues from diverse sources inside and outside the aircraft's cockpit (divided attention)" (p. 8). In addition, pilots must also maintain this monitoring of visual and auditory input over extended time (sustained attention). In a recent review of the research on aging and attentional processes, McDowd and Birren (1990) summarize the current findings related to age differences in these four aspects of attention.

In selective attention, an individual must select and focus on a single input and filter out irrelevant information. Studies of age differences in selective attention, which usually involve a visual search paradigm, have consistently concluded that older adults compared to younger adults are more distracted by irrelevant information (Plude & Hoyer, 1985). Specifically, older adults appear to have more difficulty discriminating relevant from irrelevant information, particularly when the task is effortful or demanding. Using their battery of aviation-relevant tests, Braune & Wickens (1985) found that performance on selective attention tasks declines with age. Boer (1986), studying a wide age range of pilots, found a strong relationship between increasing age and poorer performance on this type of attention task. Further support for the important role of selective attention in pilot proficiency comes from research by Gopher and Kahneman (1971) and North and Gopher (1976), who found that selective attention tasks were significant predictors of flight proficiency. Glanser and Glaser (1959) studied Air National Guard and commercial airline pilots (who ranged in age from 20 to 50) and found that older pilots were less successful than younger pilots at detecting relevant from irrelevant targets in an aviation-relevant display. Thus, selective attention appears to be an important predictor of flight performance and age decrements in selective attention in aging pilots have been found. However, practice may reduce these age effects. Hoyer (1987) has found that expertise can reduce these age-related declines in selective attention.

Particularly during high workload conditions, pilots need to perform multiple tasks at the same time and must be able to divide attention between different inputs and activities (Braune & Wickens, 1984a; North & Gopher, 1976). Divided attention, "the process by which attention is controlled to successfully perform two simultaneous tasks" (McDowd & Birren, 1990, p.223), appears to be characterized by an age-related deficit. However, there are some data inconsistent with this general conclusion. Specifically, some studies (Salthouse & Somberg, 1982; Wickens, Braune, & Stokes, 1987) have found no age differences in divided attention performance. It should be noted that the tasks used in these studies were simple perceptual and memory search

tasks. With more complex tasks age differences in divided attention performance are found. For example, Ponds, Brouwer, and van Wolffelaar (1988) found significant age-related decrements in performance on a simulated driving task. (Subjects over age 60 performed more poorly than those under age 60). This supports Hunt and Hertzog's (1981) proposal that attentional capabilities (such as divided attention performance) decline with increasing age. Complex tasks that demand more attention will show larger age-related deficits than simple tasks. Thus, it appears that the bulk of the evidence suggests an age-related deficit in divided attention that is particularly manifested on more complex tasks. However, a crucial question still remains about the role of practice and expertise in mitigating this age-related decline in divided attention performance.

The available literature in the area of age differences in attention switching (the process of alternately monitoring more than one source of information) is also characterized by theoretical controversy and contradictory findings. Many of the available studies are aimed at examining the relationship between attention switching and short-term memory. Some researchers (Welford, 1964) have proposed that attention switching reduces short-term memory capacity and that this reduction is most pronounced in older adults. Other researchers (McGhie, 1969; Schonfield, 1981) propose that age-related differences in memory retrieval are responsible for older adults' problems with attention switching tasks. Neither of these theoretical positions has been consistently supported in experimental studies. The nature of the relationship between short-term memory and attention switching is still unclear. More important to note is the fact that age-related differences in this aspect of attention have not been consistently reported. McDowd and Birren (1990) reviewed the few available studies and concluded that visual attention switching does not decline with age but that auditory attention switching does. For example, using an attention switching version of a dichotic listening task, Braune and Wickens (1985) did find that attention switching declines with age (subjects were 20 to 60 years old).

Sustained attention, usually assessed with a vigilance paradigm, has not received much research attention in the area of cognitive aging. The few studies which have been done indicate that on vigilance tasks older adults (over age 70) are less accurate in detecting the target stimulus than young adults (under age 30) and that this age difference remains even after extended practice (Parasuraman & Giambra, 1991). Most researchers explain this age-related difference in detection accuracy on vigilance tasks as due to age-related differences in central nervous system arousal, however they do not agree on whether older adults are overaroused or underaroused in comparison to younger adults (McDowd & Birren, 1990). While most studies find a main effect for age on vigilance tasks (with older adults performing more poorly than young adults), age does not seem to interact with the length of the task. "The vigilance decrement that normally occurs during an extended task does not seem to be greater for older than for young adults" (McDowd & Birren, 1990, p. 225). Since pilots must maintain performance over extended periods of time, sustained attention is a critical component of flying. Therefore, these age differences in detection accuracy in laboratory-based vigilance tasks have important implications for pilot performance and age-related differences in pilots' performance on vigilance tasks should be investigated further.

Memory. Much of a pilot's activity requires memory operations. Salthouse (1985) reviewed the literature on aging effects on memory and concluded that age is responsible for between 4% and 24% of the variance on various memory tasks. It is clear that increasing age

14

differentially affects different memory processes. Hultsch and Dixon (1990) draw the following conclusions about those memory systems that show changes with age and those that are spared. With regard to Craik's (1977) well-known distinction between primary (short-term) memory and secondary (long-term) memory, there still appears to be support for the notion that age effects on primary memory are minimal, unless these short-term memory tasks also require active manipulation of information or division of attention. There are significant age-related deficits on tasks involving retrieval from secondary memory that may also reflect age differences in initial encoding processes. Hultsch and Dixon (1990) also note that working memory, which requires the simultaneous storage of recently presented material and processing of additional material, is a critical system that may show substantial age-related decrements. Older adults' deficits appear to lie primarily in the processing rather than storage aspects of working memory (Hultsch & Dixon, 1990). These age-related declines in working memory are important because of working memory's contribution to performance on other cognitive tasks (Salthouse, 1985). Salthouse (1991) proposes that these age-related differences in working memory may be mediated by age-related decrements in processing speed. However, generalization from age differences in laboratory performance to the performance of aging pilots are suspect, because working memory appears to be domain specific and influenced by experience (Charness, 1989). The crucial question is: to what degree does poorer working memory influence pilots' performance of a well-practiced complex task.

In the aviation literature there are few studies that would allow conclusions to be drawn about age-related differences in memory performance of pilots and the relationship of these differences to pilot performance. While Braune and Wickens (1984a) found that performance on Sternberg memory and delayed digit recall tasks (as measured with their aviation skills battery) declined with age in non-pilot subjects (aged 20 to 60), they (1984b) found no age-related declines on these measures in instrument-rated pilots (aged 20 to 60). Further, while showing no age decrement in pilots, the Sternberg memory task was found to be a strong predictor of performance on a GAT-2 twin engine simulator (1984b). Thus, while there appears to be support for the conclusion that memory performance does influence pilot performance, it is not clear to what degree pilots (as compared to the general population) exhibit performance decline on these relevant memory tasks and what role expertise and experience may have in minimizing age-related differences on domain-specific memory tasks.

Information processing speed. It has been well established that speed of information processing decreases with advancing age (Cerella, 1990; Kline & Szafran, 1975; Kline & Birren, 1975). Because much of a pilot's behavior in high workload situations must be rapid as well as accurate (Braune & Wickens, 1984a), it is important to consider to what degree this slowing in processing speed is found in pilots and to what degree it influences flight proficiency. Wickens, Braune, Stokes, and Strayer (1985) studied non-pilot adults between age 20 and 60 looking for age-related performance differences on a revised version of the Braune and Wickens (1984a) aviation-relevant skills battery and found age-related slowing in processing speed which was mostly localized in response processing. While this age-related slowing has been found in healthy, educated subjects, it is not clear to what degree it occurs specifically in pilots (Institute of Medicine, 1981). Even if pilots are found to exhibit the same magnitude of slowing on laboratory-based information processing tasks that has been found in other healthy older adults, further research will be needed to determine the impact of this slowing on flight performance.

15

Judgment and decision making. There is mounting evidence that errors in judgment and decision making contribute to a large proportion of aviation accidents and incidents (Jensen & Benel, 1977). Jensen (1982) estimates that faulty pilot judgment is involved in about 50% of civil aircraft accidents. With the increasing automation of the cockpit, the pilot's role as a decision maker appears to be becoming even more prominent. However, most of the data relevant to this aspect of pilot performance is based on post-hoc analysis of accident reports. There is currently very little experimental data related to pilot decision making (Wickens & Flach, 1988) and even less known about the effects of aging on this aspect of pilot behavior. The research that is being conducted on decision-making processes in commercial aviation is focusing heavily on crew decision making rather than individual decision making. The one study (Szafran, 1969) to examine age-related differences in pilot decision making found no significant age effects.

While age per se has not been found to influence pilot decision making, there is ample speculation that the increased level of experience that characterizes older pilots does. Mohler (1981) notes that there are numerous anecdotal accounts of older healthy captains utilizing their well-developed judgment and decision making to avert an accident. If the level of experience and expertise of the pilot does have an effect on decision making (Mosier, 1991), the question becomes how is the experienced pilot's decision making different from the less experienced pilot's? Before describing a set of studies that have addressed this question directly, it is important to consider on a general level what differentiates expert from novice decision makers.

Klein (1989) has developed a model of expert decision making, termed Recognition Primed Decisions. There are two critical phases in this decision making model: the assessment of an uncertain problem situation and the systematic evaluation of plausible problem solutions. According to this model, expert decision makers are able on the basis of their prior experience to quickly recognize and appropriately classify situations and to identify the most plausible course of action. When experts perform highly practiced tasks in familiar operational situations, they recognize and respond to complex situations using pattern recognition or intuition (Chappell, 1991). Experts also frequently immediately recognize the best solution to a problem without evaluating alternatives (Klein, 1989).

Stokes and Raby (1989) suggest that two factors appear to exert an influence on pilot decision making: the experience level of the pilot and the conditions of stress in which the decision is carried out. These two factors are being investigated at the University of Illinois Institute of Aviation. Stokes and his colleagues (Wickens, Stokes, Barnett, & Davis, 1988) have used a computerized pilot decision-making simulator (MIDIS) to investigate an information processing model of pilot decision making that takes into account pilot experience and stress. They found that low and high experienced instrument-rated pilots did not differ significantly in their judgments, but that experienced pilots were more confident of their decisions. "The two groups however did differ in terms of what problem variables degraded decision performance, and what individual abilities affected that performance. In particular, novice decision performance was partially predicted by information processing tests related to spatial abilities, working memory capacity, mathematical ability and by tests of declarative knowledge. However, these tests had little predictive abilities for the more experienced pilots" (p. iii). In a related study (Stokes, Belger, & Zhang, 1990), novice pilots were found to show significantly

poorer decision making when under laboratory-induced stress. This stress did not lead to poorer decision making in the experienced pilots.

Clearly, judgment and decision making are important aspects of pilot performance that deserve increased research attention. In particular, the research has hardly begun to address the effects of aging on this aspect of pilot performance. The research on novice versus experienced pilots is intriguing but requires further examination.

2.3.2 Other Pilot Characteristics

Since the present project is aimed at examining the influence of aging on pilot performance, this section of the review has focused primarily on the role of pilot age. However, as can be seen in the conceptual model presented in Figure 1, other pilot characteristics also exert critical influences on pilot performance. Two of the most important, pilot health status and level of experience, will be addressed in section 3 of this report but will also be summarized briefly below.

Presence/absence of disease. When considering pilot characteristics as predictors of pilot performance, it is critical to consider health status as well as age to disentangle as much as possible the effects of primary and secondary aging. For example, Szafran (1966; 1969) studied the performance of older pilots on a number of psychological and psycho-physiological tasks and found that age was not a significant predictor of performance, but cardio-pulmonary status was. Review of the aviation medical literature in section 3 of this report describes additional studies that have linked medical variables, age, and pilot performance.

Experience. Pilot experience and expertise is also a critical characteristic that influences pilot performance. The literature related to the influences of pilot age, recent flight time and total flight time on aviation accidents (described in section 3 of this report) strongly demonstrates the influence of experience on pilot performance. In particular, the role of recent experience appears to be critical. Childs and Spears (1986) reviewed much of the literature relevant to the role of experience in pilot performance. Unfortunately, most of the literature they surveyed is in the form of relatively inaccessible reports. They conclude that "pilots experience significant performance decrements over short periods of time ... and such ... decrements are mainly cognitive in nature (Childs & Spears, 1986, p. 237). "That the decrement in performance occurs over a short period of time seems to suggest that age, itself, is not an important factor in performance. Golaszewski's (1983) findings, to the extent that they can be trusted, bear on this question. He concluded that older pilots with little recent flight experience have a high accident rate. It may be that older pilots have greater difficulty reacquiring the skills that are lost over the short term when pilots do not maintain currency.

2.3.3 Stressors: Their Interaction with Age

Stress. Green (1985) identifies three kinds of stress which affect pilots: environmental stress (noise, vibration, cold, etc.), acute reactive stress (situations in which "fight or flight" occurs), and life stress (stress produced by bereavement, divorce, etc.). Upon examining the records of approximately 300 aviation accidents and incidents in the U.K. Confidential Human Factors Incident Reporting Programme, he concluded that with the exception of sleep

17

deprivation, the importance of which seems to be underestimated in analysis of accidents, environmental stress is mild and is a relatively unimportant factor in accidents. The meager experimental evidence does indicate that acute reactive stress can be disruptive. Analysis of some aviation accidents suggests that in some cases aircraft emergencies can increase a pilot's arousal level and contribute to pilot errors. Green also concludes that life stress may be an important factor in causing accidents. However, most of the available data is anecdotal and consists of the retrospective accounts of pilots who have been involved in accidents and incidents. For example, Ursano (1980) and Green (1985) cite case studies to illustrate the ways in which life stresses, such as marital difficulties and interpersonal problems with work supervisors, can be implicated in accidents. Alkov, Gaynor, and Borowsky (1985) employ a more epidemiological approach and provide evidence that life stress may be an important factor in aircraft accidents. Naval aviators who had been involved in major aircraft mishaps were divided into two groups, those who were and were not causally involved in their mishaps. The group identified as culpable for their accidents reported higher levels of recent stressful life events. We are unaware of any literature that describes the relationship between age of the pilot and stress.

Drugs. Studies of drug effects on pilots focus mainly on alcohol and marijuana. Modell and Mountz (1990) survey the problem of drinking and flying, pointing out its special dangers and its implication in general aviation accidents. They note that approximately 10% of general-aviation pilots involved in fatal aircraft accidents due to pilot error had measurable blood alcohol concentrations. They further note that while alcohol is often implicated in general aviation accidents "no fatal accidents have been recorded within the U.S. airline-transport system in which alcohol was found to be a causative factor (Modell & Mountz, 1990, p 455)."

Direct studies of the effects of alcohol on performance (Collins & Mertens, 1988; Morrow, Leirer, & Yesavage, 1990; Ross & Mundt, 1988) show that alcohol deleteriously affects numerous aspects of pilot performance, including psychomotor and spatial abilities, radio communication, localizer tracking performance, and collision avoidance. Collins and Mertens (1988) and Morrow, Leirer, and Yesavage (1990) found that older pilots are more impaired by alcohol than younger pilots.

Yesavage, Leirer, Denari, and Hollister (1985) found that marijuana ingestion impaired pilot performance on a simulated landing task. Leirer, Yesavage, and Morrow (1989) subsequently showed that marijuana ingestion impairs the performance of older pilots to a greater extent than it does younger pilots, particularly on pilot tasks involving working memory.

High workload. The research of Szafran (1966; 1969) has demonstrated that workload is a critical factor to consider when examining age effects on pilot performance. Mertens, Higgens, and McKenzie (1983) found a significant interaction between age and workload on a battery of aviation-relevant cognitive tasks (the Multiple Task Performance Battery). These tasks included measures of tracking, problem solving, visual target identification, and mental arithmetic. Workload was varied by requiring time-shared performance of different combinations of these tasks. At all age levels, increasing workload caused decrements in performance but the amount of decrease was greater as age increased. These findings suggest that there may be age-related decrements in important flight related functions that would adversely affect performance but these decrements would only be revealed under moderate or

18

high workload level. In a simulated operational situation, they suggest that a secondary monitoring task may be useful in this regard.

2.3.4 Assessing Pilot Performance

The model presented in Figure 1 is aimed at conceptualizing the relationship between pilot age and performance. A critical component of this model is the proposal that pilot performance be assessed at three levels of abilities:

- Domain-independent psychomotor, perceptual, and cognitive skills which are not specifically related to flying but are assumed to influence pilot performance.

- Domain-dependent knowledge, judgment, and decision making related to piloting an aircraft.

- Complex pilot performance under varying flight conditions.

The remainder of this section will review the available literature relevant to these three aspects of pilot performance.

Domain-independent cognitive skills. There is a large literature on pilot selection and training, particularly in military contexts, that attempts to identify the perceptual, psychomotor, and cognitive skills presumed to be important to piloting performance. Many of these research programs have involved the construction of aviation-relevant cognitive skills batteries that are then validated against some criterion, most frequently success in flight training. Some of the most widely used batteries are the:

(1) Basic Attributes Test Battery (BAT) (Carretta, 1987a).
(2) Complex Cognitive Assessment Battery (CCAB) (Samet, Geiselman, Zajaczkowski, & Marshall-Miles, 1986).
(3) Microcomputerized Personnel Aptitude Tester (MICROPAT) (Bartram, 1986).
(4) Naval Aviation Selection Test Battery (Damos & Gibb, 1986).
(5) Simple Portable Aviation-Relevant Test-Battery and Answer-scoring System (SPARTANS) (Stokes, Banich, & Elledge, 1988).
(6) Unified Tri-Services Cognitive Performance Assessment Battery (UTC-PAB) (Perez, Masculine, Ramsey, & Urban, 1987).
(7) COGSCREEN (Horst & Kay, 1991).

Complete descriptions of each of these batteries can be found in Appendix A, as well as more detailed analysis of each battery's reliability and validity.

Since much of this research focuses on the construction of pilot selection test batteries (e.g., BAT, MICROPAT), the results focus on the relationship between performance on the test battery and successful completion of flight training. There is much less data available concerning the relationship between performance on the battery and actual pilot performance, particularly in skilled pilots. Other batteries have been constructed to examine aviation-relevant

19

cognitive skills, but not for the purpose of pilot selection. These batteries (e.g., COGSCREEN, SPARTANS) are aimed at assessing the effects of aging or neurological impairment on these aviation-relevant skills. While not specifically constructed for pilots, there are also additional batteries that assess similar types of cognitive skills (e.g., CCAB, UTC-PAB).

This section of the report will summarize information available about the domain-independent cognitive skills assessed by these batteries and the degree to which they may be useful for detecting decrements in the cognitive performance of commercial airline pilots occurring from the effects of aging or other impairments. This summary will also focus on the utility of the test batteries as measures of pilot performance and specifically, the feasibility of using these batteries to measure the proficiency of pilots over age 60.

Test development in aviation has been a long and difficult process. Guilford (1947) attributed the failure to predict success in flight training to three factors: (1) the tests were not designed to predict flight performance, (2) motivational factors compensated for weakness in personality traits during training, and (3) subject bias yielded inaccurate measures of the personality trait under study. Some of these problems remain today, but the problem of differentiating among experienced pilots is an even more daunting task. Identifying cognitive deficits in experienced pilots is a conceptually different enterprise from selecting candidates for flying training from an untrained population.

Most attempts to improve the predictive validity of aptitude test batteries have some inherent methodological problems. The test variables have often been related to global criterion performance measures in training such as graduation/elimination or composite flight grades. These performance criteria may obscure the components of skilled performance or behavioral attributes associated with the selected sub-tests. Helmreich et al. (1986) pointed out that different combinations of predictors relate to quite different measures of performance at different points in time.

Tests developed and validated to select candidates for flying training should not be used to measure performance of highly trained and skilled commercial pilots, without further validation on the target population. There is often little correspondence between tests that have high validity for training criteria and those that predict success on the job. For example, tests of cognitive ability that have been proven as good predictors of the ability to learn are not comparably good measures of skilled performance.

The test population itself poses a methodological problem in validation. Most selection batteries are developed for testing heterogeneous groups. Airline pilots are a very homogeneous population that severely restricts group variance and the ability of the tests to discriminate among subjects.

Keeping these methodological issues in mind, there are some general conclusions that result from the review of these seven cognitive skills batteries. The purpose for the development of the test batteries was to either identify the least competency required for successfully completing basic flying training or to identify decrements in certain cognitive skills. The tests have research support for their utility to accomplish their intended purposes.

However little has been found to show that the tests can be extended to reliably discriminate among levels of performance for experienced pilots. None of the test batteries reviewed have been validated against the performance of experienced pilots nor have they been normed for pilots over age 60. Little aviation-specific normative data is presently available. These two steps need to be accomplished before it can be determined if the batteries are appropriate for determining an experienced pilots' fitness to fly.

The sensitivity, specificity and utility of some of the batteries has been suggested from past research attempting to discriminate between experienced pilots and clinically impaired patients. However, this review did not find the batteries performing at a level necessary to reliably classify normal populations according to levels of cognitive functioning.

Some of the batteries were intended to be used as adjuncts to existing selection procedures. This still appears to be the most promising approach to using the batteries in discriminating among experienced pilots. If a point-to-point theory has merit here, then the more realistic computer-assisted approach should have more predictive power than earlier cognitive test batteries. Point-to-point theory states that the more common points between the predictor and the criterion space, the higher the validity coefficient. The combination of computer-based cognitive skills tests with flight simulators and other relevant selection methods would be appropriate. It is also suggested that if a cognitive test battery is to be used the sub-tests should be selected on the basis of: (a) is the sub-test measuring a skill that is affected by aging, (b) is the skill being measured critical to the task, and (c) can the skill be compensated for by experience and training.

<u>Domain-dependent pilot knowledge. judgment and decision making.</u> The types of pilot selection and evaluation batteries described in the last section primarily assess domain-independent cognitive skills. As Banich, Stokes, and Karol (1990) conclude, these batteries assess the kinds of skills most likely to be detrimentally affected by neuropsychological impairment or aging. These batteries were not designed to and <u>do not</u> measure pilot expertise. Comprehensive evaluation of pilot competence would not be complete without including some measures sensitive to the advantageous effects of increasing expertise on pilot performance. In particular, it appears necessary to assess aspects of pilot knowledge, judgment, and decision making that may not be detrimentally affected by aging and may even improve with increasing expertise. As Mohler (1981) notes, pilot judgment and reasoning tend to be preserved in older pilots and may compensate for some of the age-related losses in domain-independent cognitive skills.

Since there has been so little research done on individual decision making in pilots, there are really no well-developed, validated measures that assess this elusive aspect of pilot expertise. The two most promising experimental measures of pilot knowledge and judgment were recently developed by Stokes (1991) and his colleagues (Stokes, Belger, & Zhang, 1990) at the University of Illinois Institute of Aviation. Each of these measures will be described below. However, it should be noted at the outset that while the measures seem to have much promise for assessing the influence of expertise on pilots' knowledge and judgment, further development is required before they can be used with commercial pilots.

Expertise can be defined as a highly developed repertoire of pattern-oriented representations (Chase & Simon, 1973). Thus, one aspect of expertise that may be important to assess is its impact on knowledge representations in long-term memory. The first measure, "FLITESCRIPT", developed at the University of Illinois Institute of Aviation, was designed to index pilots' representations of situational knowledge in long-term memory. It is analogous to the well-known chess experiments conducted by DeGroot (1965) and Chase and Simon (1973). There are two versions of the FLITESCRIPT test, a recall version and a recognition version. The recall version of the test involves reconstructing both randomized and coherent air traffic control (ATC) radio call sequences from memory. The recognition version requires listening to an ATC communication sequence and selecting the correct graphic depiction of the situation represented by the ATC communications from a set of alternatives. "By controlling out simple memory effects by reference to recall of the random sequences, the quality of reconstruction, we believe, must be influenced predominantly by the availability of appropriate situational 'scripts' in the pilot's repertoire" (Stokes, Belger, & Zhang, 1990, p 19). These scripts are presumed to be present to a higher extent in experts than novices. Stokes' research has shown that this knowledge representation task is a better predictor of expert pilot performance than are cognitive skills tests (Stokes, Belger, & Zhang, 1990). While modifications will need to be made in the test, including some revisions of the scoring system, this appears to be a promising measure to assess a crucial aspect of pilot knowledge that may be affected by the development of expertise.

Stokes and his colleagues have also developed a microcomputer-based simulator for cockpit decision making training and research called "MIDIS". MIDIS was developed to investigate "pilot judgment in uncertain high-risk multi-cue decision-making contexts" (Stokes, 1991, p 109). It is implemented on an IBM PC/AT and has a full, high-fidelity instrument panel based on that of a Beech C23, a single engine fixed-gear aircraft. "The formal structure of MIDIS software is that of a 'graph traverser', in which the nodes of a branching structure are linked by transitional probabilities. Each node is a realistic in-flight situation. The system is thus capable of considerable flexibility and is able to simulate the chain of deteriorating circumstances which typifies the development of mishaps in time-critical decision-making contexts" (Stokes, 1991, p 107). At various points during the simulated flight, "subjects are allowed to enter what they consider to be the optimum action alternative and indicate their confidence in that alternative. The computer scores the correctness of the decision, the confidence rating, and the latency within which subjects chose the alternative" (Stokes, et al., 1991, p 314).

There are several attractive aspects of MIDIS as a measure of individual pilot decision making. First, it presents problems, embedded in a coherent flight mission and thus has high face validity. Second, the types of flight scenarios programmed into MIDIS span the breadth of different cognitive and attentional demands (Stokes, et al., 1991). Third, the MIDIS system permits testing under differing conditions of workload and stress. Thus, MIDIS presents exciting possibilities for the examination of the effects of aging and experience on pilot judgment in particular and pilot performance in general. However, further development of the system with respect to scenario development and decision rating is needed. The question of whether a system configured for a single engine aircraft is appropriate for commercial pilots also needs to be addressed.

Sarter and Woods (1991) review one other area of aviation research relevant to the issue of pilot expertise based on knowledge, judgment and decision making. This body of research has been loosely labeled "situation awareness". While this phenomenon is frequently cited in the aviation psychology literature as a critical aspect of pilot performance, there is no commonly accepted definition or agreed upon methodological approach for measuring it. In its broadest sense situation awareness appears to be the pilot's complete, coherently organized, flight-relevant knowledge that is continually updated over time. It is the ability to not only assess the aircraft's current situation but also to project its future status (i.e., "being ahead of the plane") (Sarter & Woods, 1991). Previous attempts to measure situation awareness have frequently involved intrusive in-flight probing techniques, such as blocking out cockpit displays at random points during a simulated mission and requiring the pilot to answer questions about the flight situation at that moment in time (Endsley, 1988).

Other studies have attempted to assess pilot situation awareness by asking the pilot questions during a post-flight debriefing session. Both of these approaches are restricted in that they focus on isolated components of situational knowledge rather than awareness of the overall situation. They also require that the pilot answer questions based on memory for the flight situation. However, this memory is recalled without the aid of the complex dynamic flight situation. In other words, the pilot might have been aware of much more information during the actual flight context than he is able to recall out of that context. Therefore, Sarter and Woods (1991) propose that to adequately assess situation awareness, it is necessary to stage complex dynamic situations in high-fidelity, full-mission simulations that require retrieving and integrating competing information from a variety of sources over time. Pilots' situation awareness in these simulations should be assessed with less intrusive and artificial probing. For example, they suggest that the experimenter might play the role of air traffic controller, asking the pilot for information that would assess the pilot's implicit knowledge of the situation. Such an approach seems to be a promising way to examine how pilot experience influences flight-relevant knowledge and judgment.

Complex pilot performance: Simulators and airline proficiency checks

Simulator utilization. High fidelity simulators have assumed a commanding role in the process by which Part 121 aircarriers train pilots, certify their competence, and conduct periodic proficiency checks. Simulators provide numerous advantages when compared to conducting training and/or certification in actual aircraft. In addition to the obvious fact that the hourly cost of operating a simulator is much less than the cost of large aircraft, other benefits include the opportunity for more in-depth training, increased safety, fuel conservation, and reduced noise pollution. The framework within which simulator training and certification is conducted is the FAA National Simulator Evaluation Program. Before a simulator can be legally used to provide training or checking, it must be certified by the FAA. FAR Section 121.407 mandates that simulators must meet and maintain strict performance and functional characteristics. Guidelines for simulator certification are contained in FAA Advisory Circular 120-40A. Devices which resemble simulators in form and function but that for various reasons do not meet established criteria for legal certification are considered to be training devices.

The utilization of simulators as part of the overall Age 60 project evaluation scheme, as discussed in the Age 60 performance evaluation model (described earlier in this review) has

several positive attributes that go beyond the scientific issues of determining complex pilot performance. Logistically, simulators are presently used by all of the Part 121 aircarriers. They are integral to the industry. Pilots are routinely scheduled to be trained and evaluated in simulators. As a consequence, pilots accept simulators and simulator results as a valid tool for assessing their abilities. It is an established part of their culture.

Line-oriented flight training. Air carriers rely heavily on simulators for their flight training and proficiency checking. The commonly used simulator training process is known as Line-Oriented Flight Training (LOFT). LOFT training uses a complete crew flying complete flight segments that contain the normal, abnormal, and emergency procedures that are representative of aircarrier flight operations. LOFT is the procedural backbone of the Part 121 aircarrier training process. All LOFT must comply with the requirements delineated in FAR Section 121.409(b). Guidelines for the design and implementation of LOFT simulations are contained in FAA advisory Circular 120-35B.

One aspect of LOFT that is of interest in the context of performance evaluation (as opposed to training) is the Line Operational Evaluation (LOE). This concept is tied to the FAA *Advanced Qualification Program (AQP)*. *AQP is the subject of the FAA proposed Special Federal Aviation Regulation (SFAR) No. 58.* LOE is designed to check both crew and individual competence. LOE differs from LOFT in a fundamental way. LOFT is "no-jeopardy" training, in that an instructor does not issue a passing or failing grade to an individual crewmember. Flawed decisions are considered as contributing to the learning process. Performance deficiencies encountered in a LOFT session are addressed with additional training. At the end of a LOFT session, the instructor certifies that the LOFT session has been completed. Line Operational Evaluation, however, contains elements of "jeopardy". That is, the instructor conducts evaluation of both crew and individual competence and performance. The development of LOE and the specific methods by which evaluations will be conducted is tied to the AQP development effort, which is still unfolding. Hence, LOE represents a near-term potential rather than a current approach for performance evaluation. AQP guidelines call for the development of scaler measures of performance by which individual performance can be quantified. Successful realization of these stated goals would be useful in the context of the needs of AGE 60 project and bear watching. The use of a LOFT scenario with unexpected situations may hold more promise as a performance assessment tool than the use of a standard check-flight scenario. However this is complicated by the difficulty of assessing individual performance in a closly integrated crew environment (National Institute on Aging, 1981).

Crew coordination evaluation. The issue of crew coordination as an influence on overall performance in the cockpit is beginning to have a major impact on air carrier flight training and evaluation. Cockpit Resource Management programs are beginning to be adapted in the industry. For example, Federal Express and Delta Airlines (and others) have CRM training programs designed to increase crew proficiency in the area of CRM. Crew coordination issues (Cockpit Resource Management) grow increasingly important with the increased role of the pilot as a system manager in the glass cockpit aircraft, such as the MD-11 and the Boeing 777. Recent work at NASA Ames has focused on the measurement of the level of crew coordination and the associated quality of decision-making in the cockpit using the NASA Boeing 727 simulator located at Ames Research Center (Chidester, 1990; Mosier, 1991). Indications from that ongoing work suggest that the quality of information transfer among the flight crew predicts

24

both process (crew coordination) as well as operational (errors) criteria. Although this research focused on measures for assessing crew performance, it appears that some of the measures are equally applicable in assessing individual performance. Thus, the NASA work should have potential for providing scenario development guidelines and performance measurement techniques that could contribute to the Age 60 project.

Performance assessment. Historically, there were two major approaches to the development of the criteria and means for the assessment of pilot performance; 1) qualitative evaluation of performance based primarily on instructor ratings and judgement, and 2) quantitative grading of performance based on numerical rating scales and recordings of pilot actions which reflect the quality of performance (Gerathewohl, 1978). Subjective ratings of pilots have inherent reliability problems. Currently, proficiency check methods used by Part 121 aircarriers continue to be subjective in nature. The individual being evaluated flys a scenario that presents routine and non-routine situations. Response is observed by an check instructor and judged to be acceptable or non-acceptable. Thus the exercise is a pass/fail evaluation. Individuals who fail portions of the check flight are given the opportunity to re-fly the evaluation, with additional instruction if necessary. Although the qualitative assessment approach is subjective in nature and is subject to possible instructor bias, it is widely used today in the industry and has proven itself to be operationally sound. The development of automated, objective measures and techniques in an aircraft or simulator will require the development of performance standards and acceptable criteria (Institute of Medicine (1981).

Simulator Training Methods Simulator training methods are well-developed and proven, however, techniques for quantifiable performance measurement in a simulator is less so. Most research along these lines is oriented to military related requirements, where performance requirements are imposed by the unique needs of combat (Kruk, Regan, Beverley, & Longridge, 1983). Although scientifically interesting, development in the military context is of little use in the Age 60 project.

Research conducted at the U.S. Air Force Human Resources Laboratory at the Wright-Patterson Air Force Base has focused considerable effort toward development and validation of aircrew performance measures (Waag, 1988). Development emphasis was placed on criterion-referenced or behavioral objectives in an attempt to standardize the measurement of aircrew performance. Ultimately, a Performance Measurement System (PMS) was developed to interface and function with the C-5A aircraft flight simulator located at Altus AFB, Oklahoma. PMS had the capability to measure specific parameters of interest and also to assess performance using a scoring algorithm. PMS was independent and autonomous to the host simulator, but interfaced to the extent necessary to passively monitor all required I/O signals, and to pass control and malfunction data to the host.

Scenario development. There is a useful base of literature on the subject of the simulator as a training device. Inseparable from the subject of simulator training is the issue of scenario development. That is, the scenario is the framework that drives the exercise. It is the key that elicits the actions and reactions that support the entire simulation event. Consequently, scenario development is a critical first step in preparing for any intended simulator exercise, be it general training, specific type training, performance evaluation of individuals, or observation of crew coordination and management. Scenario development must reflect training objectives and must

elicit desired results. Feedback from students to instructors and scenario developers is important (Gregorich, 1991). An exhaustive analysis of simulator requirements, piloting tasks, and task performance requirements was conducted under FAA sponsorship (Guilliom et al, 1985). This analysis developed an overall approach for simulator based training and checking. The resulting approach was termed Airman Certification System Development (ACSD). Of particular interest is the delineation within ACSD of Type A objectives and Type B objectives. These are defined as Type A objectives related to proficiency in fine motor control of the airplane and its systems or Type B objectives related to proficiency in cognitive-procedural aspects of performance during which only basic control skills are required.

2.4 Conclusions and Recommendations

In 1981, the National Institute on Aging panel was not able to identify a medical or performance appraisal system that could identify those pilots who would pose a safety hazard because of early or impending deterioration in health or performance, and therefore they recommended that the Age 60 Rule be retained. However, they also made two major recommendations:

- that more study and development be carried out to develop quantifiable, objective criteria for measuring pilot performance.

- that systematic collection of performance data needed to further evaluate the Age 60 Rule be carried out.

Ten years later, despite increased research attention having been paid to these issues, these two recommendations have not been sufficiently implemented. There remains a pressing need for a comprehensive, objective system for evaluating pilot performance. Once such a system has been developed, it is also critical that longitudinal analysis of the effects of aging on pilot performance be carried out. These are the major goals of Tasks 4 and 5 of the present project. Task 4 is aimed at development of a methodology to quantitatively and objectively assess an individual pilot's performance. Task 5 is aimed at planning a longitudinal examination of the effects of aging on pilot performance. The information obtained from this project should aid understanding about the effects of increasing age and experience on pilot performance. It should also provide additional information about the feasibility, reliability, and validity of this type of individual performance assessment as an alternative to the Age 60 Rule. However, a definitive stance on the Age 60 Rule will also require consideration of pertinent medical, economic, and political issues beyond the scope of the present tasks.

2.4.1 Criteria for Developing and Evaluating a Performance Assessment Methodology

A performance assessment measure of pilot proficiency that could eventually replace the existing Age 60 Rule should not only be judged against the standard psychometric criteria of reliability, sensitivity, specificity, and both construct and criterion validity, but it should also be developed in response to the following criteria:

- Is it measuring a skill that is affected by aging?
- Is it measuring a skill that is affected by experience?

26

- Is the skill being measured critical to the task of flying an aircraft?

Specifically, several measures at each of these three levels of performance should be investigated:

- Domain-independent psychomotor, perceptual, and cognitive skills which are not specifically related to flying but are assumed to influence pilot performance. These are the types of skills which are most likely to be detrimentally affected by aging or neurological impairment.

- Domain-dependent knowledge, judgment, and decision making related to piloting an aircraft. These are abilities that may be enhanced as a result of increased experience.

- Complex pilot performance in a simulator under varying flight conditions.

In order to insure aviation safety, pilot proficiency should also be tested under a range of flying conditions. Performance of complex maneuvers under stress or in novel situations is more likely to be impaired in older pilots than well-learned familiar tasks (National Institute on Aging, 1981). For example, age-related declines in basic skills could be compensated for by increases in pilot knowledge and judgment that occur as pilots' expertise increases. However, increased knowledge may only compensate for declining skills under ideal or routine task conditions. There may be age-related decrements in important flight-related functions that would adversely affect performance that would only be revealed under moderate or high workload level.

2.4.2 Longitudinal Examination of Aging and Pilot Performance

This section of the report reviewed the available scientific literature bearing on the influence of aging on pilot performance. Unfortunately, quantitative, objective data on the flight performance of pilots varying in age do not exist. The available data on laboratory-based performance tests of basic perceptual, psychomotor, and cognitive skills suggests that there are statistical decrements in performance with age. However, while average performance on these performance measures decreases with age, variability in performance increases with increasing age. Most importantly, these laboratory-based performance tests have not yet been demonstrated to be predictors of piloting performance.

There are also very few studies that have examined the effects of aging on the performance of complex professional skills and even fewer studies that examine the contribution of practice or experience to the performance of these complex skills. However, the available literature suggests that, on some complex tasks, older adults perform as well as younger adults even though the older adults exhibit age-related declines on the component psychological skills that contribute to the complex behavior on which they are successful. In other words, there appears to be compensation for declines in the basic skills. Of course this compensation is only possible when the older adult has developed expertise in the complex task. For example, it is possible that domain-dependent knowledge and judgment related to flying may compensate for declines in the older pilot's processing speed and working memory.

27

Because of the relatively small number of studies that have directly investigated the effects of aging on specific aspects of pilot performance, there is a great need for systematic investigation of multiple aspects of pilot performance across a wide age span. In order for this investigation to be most fruitful, it should be guided by a coherent conceptual framework. The framework proposed for the present project is a model of aging and pilot performance that is similar to the decrement with compensation models that have been developed in the area of cognitive aging and that have successfully contributed to understanding the concept of expertise. Key questions that should be addressed are:

- With increasing age are there declines in performance on measuresof basic skills, pilot knowledge and decision making, or simulated flying performance? At what age do the declines occur? How much individual variation characterizes these age changes?

- Are there increments in performance on any of the above measures associated either with pilots' increasing age or experience?

- Are declines in some aspects of pilot performance being compensated for by increments in other aspects of pilot performance?

In addition, the little available data on the effects of aging on pilot performance is predominantly cross-sectional. That is, the studies have compared (at a single point in time) groups of pilots that vary in age, rather than repeatedly testing the same group of pilots as they age (a longitudinal design). While there are interpretation difficulties with both of these simple research designs, a reliance on cross-sectional data is particularly difficult to interpret because different age cohorts of pilots can vary on a number of dimensions besides age (such as health, education, and experience). Therefore, future investigations of the effects of aging on pilot performance should include a longitudinal component.

The National Institute on Aging (1981) report on the experienced pilot includes a proposal for a sequential longitudinal data collection strategy that could produce information related to aging and pilot proficiency that could serve as a basis for modifying the Age 60 Rule. This strategy assumes that state-of-the-art medical and performance testing procedures have been validated. This strategy also assumes that pilots who participate in the program would be allowed to fly after age 60 under very selected conditions and with strict monitoring. The data collection strategy could be implemented without this condition (studying pilots only up to age 60), but this would limit the conclusions that could serve as a basis for modifying the Age 60 Rule. Key components of this strategy are as follows:

(1)   Pilots who wish to continue flying beyond age 59 would indicate this preference at age 55. These pilots would then undergo four annual comprehensive medical and performance evaluations and pilot proficiency tests prior to age 60 and on an annual basis after that.

(2)   This annual comprehensive testing would also be conducted on random samples of volunteer younger pilots (age range 40 to 55). These multiple cohort groups would allow for baseline performance data for younger pilots and a comparison

28

for the older group. Performance measurements collected on younger pilots as a component of this research study would not be used to determine current eligibility to continue flying.

If this approach was implemented, useful longitudinal data would accumulate within a decade. This data would be useful in detecting population trends among pilots as a group. It would also determine whether it is feasible to use an individual pilot's performance trends over time to predict that pilot's future performance ability. This would provide a sound basis on which to draw conclusions about the influence of aging on pilot performance and could serve as a basis for a decision on modification of the Age 60 Rule.

3. THE INFLUENCE OF AGE, TOTAL FLIGHT TIME, RECENT FLIGHT TIME ON PILOT ACCIDENT RATES

3.1 Specific Aims

Part II of the report will discuss attempts to relate parameters such as age and flight time to the probability of an accident. The available studies (Golaszweski, 1983; Office of Technology Report, 1990; Guide & Gibson, 1991) primarily use information from the NTSB accident and FAA medical databases. The most ambitious and most cited of these studies is that of Golaszewski (1983). First, the design and results of the Golaszewski study will be reviewed, paying particular attention to the methodology and areas for improvement. Part of the Age 60 project effort involves the development of a Consolidated Database (CDB) that incorporates information from several historic NTSB and FAA databases (including the NTSB accident and FAA medical databases) that record information about pilot performance and age. The CDB provides relationships among the databases that were unavailable to Golaszewski. Second, ways to take advantage of the CDB in the improvement of Golaszewski's design will be discussed. Finally, the literature pertinent to the relationship between accidents and medical condition will be reviewed along with ways in which the CDB can be used to examine these issues.

3.2 Background Literature

3.2.1 *Description of Golaszewski Study*

In 1983 Richard Golaszewski, working for Acumenics Research and Technology, Inc., submitted the paper "The influence of total fight time, recent flight time and age on pilot accident rates" to the Transportation Systems Center and the FAA. Golaszewski aggregated data for the years 1976 to 1980 from two databases, the "Medical History File" (FAA Medical Certification) and the "National Transportation Safety Board (NTSB) Accident".

The NTSB Accident Records database consists of data from detailed reports of all aviation accidents investigated by the NTSB and (some) self-reported aviation incidents. The CAIS medical database consists of data collected from medical examinations of all licensed pilots who take the examinations for medical certificates that fall into three classes:

- A Class I medical certificate which must be renewed every six months and is required of all air transport pilots;

29

- A Class II certificate which must be renewed yearly and is required of all commercial pilots; and

- A Class III medical certificate which must be renewed every 2 years and is require of all private and student pilots.

Note that many employers of commercial pilots require their pilots to have Class I licenses (for example, according to Table IV.N on page 39 of the 1990 Aeromedical Certification Statistical Handbook, only 41% (59,259 of 145,599) first class airmen are employed by all airlines). In addition, a pilot may use an out-of-date certificate in place of one from a lower class, provided the certificate is not out-of-date for the lower class.

The two databases give the age of the pilots, "recent" flight time, and total flight time. Golaszewski combined the information from the two databases to estimate the accident rate as a function of the factorial combination of age, recent flight, and total flight time. The data were separately analyzed for two distinct groups -- all pilots and only Class III pilots. The information utilized in the study are described in more detail below.

Golaszewski used the following categories:

- Age: 17-19, 20-29, 30-39, 40-49, 50-59, 60-69, and 70 & over.

- Total flight time: 0-100 hours, 101-500 hours, 501-1000 hours, 1001-5000 hours, 5001 hours & over.

- Recent flight time: 0-20 hours, 21-50 hours, 51-100 hours, 101-400 hours, 401 hours & over.

In general, one can calculate accident rates (per 100,000 hours of flight) for the various categories by computing the number of accidents and the number of hours flown in each category. Golaszewski obtained the age, recent flight time (in the last 90 days), and total flight time for all makes and models of airplanes for the pilots involved in accidents from the NTSB accident database. He estimated the exposure (hours of flight) of the various groups of pilots by examining the Medical History File. This file contained the medical reports of all licensed pilots. Reports for Class I pilots are filed every six months, for Class II pilots every year, and for Class III fliers every 2 years. These reports include information on total hours flown and recent hours flown (in the last six months), and the date of birth and date of medical examination for calculating age.

Golaszewski defined "recent flight time" as the amount of time flown by the pilot in the previous 12 months. The NTSB Accident database records the pilot's flight time in the previous 90 days. Recent flight time for these pilots was estimated by multiplying this number by 4. The CAIS medical database records the number of hours flown in the preceding six months. Golaszewski "normalized" the recent flight hours by taking into account the frequency with which pilots in a given class are examined and using the factors shown below:

30

Class I   -- 1
Class II -- 2
Class III -- 4

Given the above data, Golaszewski plotted accident rates in various ways. The graphs he presented are included in Appendix B. After visual examination of these graphs, Golaszewski reached a number of conclusions, summarized below: (see pp 23-24 of Golaszewski, 1983).

(1)   Accident rates are inversely proportional to recent flight time.

(2)   For low amounts of recent flight time, older pilots have more accidents than younger pilots, and the converse is true for higher amounts of recent flight time.

(3)   When recent flight time is less than 50 hours, Class I and II pilots have more accidents than Class III pilots.

(4)   Accident rates are inversely proportional to total flight time.

(5)   Independent of total flight time, Class III pilots have more accidents.

(6)   Among Class I and II pilots with 101 to 5000 total hours of flight, older pilots have more accidents than younger ones.

(7)   The older the Class III pilot, the less likely an accident until age 60, when accidents become more likely.

(8)   Pilots with over 1000 total flight hours and less than 50 recent flight hours have the most accidents, while pilots with over 1000 total hours and more than 50 recent hours have the fewest accidents.

(9)   For Class III pilots with low total flight time, if they have fewer than 50 recent flight hours then accident rates increase with age, and the converse is true if they have more than 50 recent flight hours.

(10)   For pilots with more than 50 recent flight hours and less than 1000 total flight hours, accident rates decrease with age, while the converse is true for pilots with less than 50 recent flight hours and less than 1000 total flight hours.

3.2.2 Drawbacks to Golaszewski Design

(A) Pilot Class is the wrong independent variable. The Age 60 Rule is part of the Part 121 FAR regulations (121.383c) which pertain to scheduled air carriers, yet Golaszewski included in the study general aviation accidents which did not involve scheduled air carriers. Although the accident rates are based on accidents other than those of scheduled air carriers, they are used to make inferences about accidents of scheduled air carriers as a function of age. To further complicate matters, the NTSB classifies accidents by kind of flight:

31

- Part 121 — scheduled air carrier

- Part 135 — commuter and air taxi, and

- Part 91 — general aviation.

The medical database classifies pilots by class (I, II, and III). These two sets of categories do not exactly correspond. For example, pilots employed by scheduled airlines must have a Class I medical certificate but not conversely. A pilot employed by an air taxi need only have a Class II medical certificate, but the pilot's employer may require the pilot to have a Class I medical certificate. As a point of fact, a private pilot can have a Class I medical certificate. The lack of exact correspondence between medical certificate and the nature of the flight undermines the relevance of the Golaszewski study data to the Age 60 Rule.

Further, because Class I pilots are involved in so few accidents, Golaszewski did not compute accident rates for Class I pilots alone. The rarity of accidents involving Class I pilots strongly suggests that the distribution of such accidents across the various categories will be different than for the groups of pilots studied by Golaszewski — Class III pilots alone and all pilots. The Part 121 pilots accumulate a substantial number of flight hours but contribute (proportionately) very few accidents. That is, they contribute substantially to the denominator of the ratio for accident rate but hardly at all to the numerator. It is likely that this contribution changes as a function of age, so that the apparent effects of age in the Golaszewski data may be an artifact of the way accident rates are computed, i.e., they may reflect differences in flight hours accumulated by Class I pilots rather than differences in the likelihood of accidents.

(B) Golaszewski included accidents by pilots whose medical certificates had lapsed. Thus, their flight hours are not recorded in the CAIS medical database, and these pilots only contribute to the numerators of the accident rates. Any systematic bias in the distribution of these pilots by age would distort the corresponding accident rates. For example, younger pilots may be more likely to have lapsed certificates, thus inflating their accident rates.

(C) Golaszewski performed no statistical analyses on his data, arguing (Golaszewski, 1983, p. 6):

> "the data used in the present study are not amenable to analysis using parametric statistical techniques. Such techniques are relevant when investigating the reliability of sample data to accurately portray normally distributed continuous population parameters. In the present case, the data used in this study represent the entire population of pilots and the entire population of aviation accidents... the most important measure of the reliability of the results presented in this report is whether the underlying data are accurate."

This mixes two arguments: (1) That the data are not normally distributed; and (2) that the data constitute the whole population.

First, there are statistical tests (e.g., the ANOVA) that assume normality but are robust with respect to mild departures from normality. Further, the data are likely to be normally

distributed. A member of any particular group of pilots has a given probability of being in an accident. Each observation from that group constitutes an independent Bernoulli trial. Because the observations are independent, the number of accidents for the given group is binomially distributed. The normal distribution is a good approximation of the binomial when the mean is larger than 3 standard deviations and at least 3 standard deviations less than the number of observations (Drake, 1967). Golaszewski's data satisfy this criterion. Thus the accident rates are approximately normally distributed.

Second, whether the data constitute the whole population depends on the inferences desired. We want to make inferences about the population of all possible flights. In particular, we want to infer the probability of an accident as a function of age in the population of all possible flights. The data collected by Golaszewski can indeed be used to make such inferences. To understand this, we consider the logic of statistical testing in more detail for the example of comparing the probabilities of an accident for two populations. We estimate the probability of an accident in each population by drawing a sample of observations from each population and computing the proportion of observations in each sample that are accidents. We then statistically test whether the two proportions differ sufficiently to conclude that it is unreasonable to assume that the probability of an accident is the same for both populations. If it is unreasonable to assume equal probabilities, then we expect a similar difference in the same direction when we repeat the experiment. If it is not unreasonable to assume equal probabilities we have no statistical evidence for predicting the outcome of a repetition of the experiment. Golaszewski presented data from an experiment which was more elaborate (and very large) but perfectly analogous to the above example. Would the same results occur when we repeat the experiment (the following year)?

For example, points 9 and 10 in the above summary of Golaszewski's conclusions suggest a triple interaction among age, recent flight time, and total flight time. Without a statistical analysis we do not know whether to expect that to reoccur.

(D) The Age 60 Rule represents a sharp cutoff, presumably reflecting some relatively dramatic change in the pilots as they approach 60 years. The ten year age categories used by Golaszewski are too large to get a clear picture of such a change. A more fine grained analysis, i.e., by year, would be more appropriate.

(E) As with any study of the effects of age, the outcome of the Golaszewski study may be peculiar to the particular age cohorts chosen. Replicating the study for a variety of cohorts should be performed to assess whether the same conclusions would be obtained.

(F) Another limitation of the Golaszewski study is that it considers the exposure to risk of accident solely on the basis of hours flown. Many sources indicate that most airplane accidents occur during takeoff or landing phase. Thus it may be more appropriate to compute the rate of accidents per number of landings rather than per number of flight hours. It is quite likely that the data are systematically biased so that older airline pilots have fewer landings per hour flown. It could be argued that the older the pilot, the higher the seniority, the more likely the pilot gets the choice of flights, the more likely the pilot chooses a long flight, and the more likely the older pilot has fewer landings per flight hour. Assuming that this argument is correct,

using hours of flight for the denominator for accident rate systematically deflates the accident rate for older pilots.

3.2.3 Other Studies of Age and Accidents

To the best of our knowledge, there are only two other major studies similar to Golaszewski's: an OTA report (1990) and a report presented at the 1991 meeting of the Human Factors Society by Guide and Gibson (1991).

The Office of Technology Assessment (OTA) report. The OTA report presents graphs of accident rates of Class I & II pilots as a function of (ten-year intervals of) age provided by Dr. Charles Billings, NASA Ames Research Center. The figures state that the data are from "Golaszewski 1983, and NTSB, 1990," but no further information is provided. Conclusions from these data are difficult to substantiate, because it is unclear how the data are calculated and no statistical analyses were performed on the data. Nevertheless, the OTA drew the following major conclusions:

- For Class I & II pilots with more than 1000 hours flight experience and more than 50 hours recent flight experience, "an age effect is present and is beneficial until 50, after which rates increase" (Figure 1).

- For Class I & II pilots "an increase in accident rates with increasing age is seen after age 39 in pilots with 501-1000 hr, after age 49 in pilots with 1001-5000 hr, and after age 59 in pilots with over 5000 hr" (Figure 2).

- For Class I & II pilots "who fly more than 100 hr/yr, increasing age (and probably total experience) and increased recent flying time both have beneficial effects. After age 60, accident rates increase even if pilots continue to fly over 400 hours per year" (Figure 3).

These conclusions do not bear directly on the Age 60 Rule since they are based upon both Class I and II pilots. To get such evidence, it would be necessary to consider only those data for pilots of Part 121 aircraft.

The Guide and Gibson report. The Guide and Gibson study provides few specific details of their design, but it is similar to that of Golaszewski. They obtained their data for the years 1982 through 1988 from four sources:

- NTSB Accident database;

- Aircraft Owners and Pilots Association Air Safety Data Base (AOPA);

- FAA Statistical Handbook on Aviation (providing the distribution of pilots by age and type of license); and

34

- COMSIS Research Corporation (providing the flight hour distribution for Class I, II, and III pilots in 5 year age groups, as compiled from FAA pilot medical records).

They used 5-year age categories starting with 20-24 and finishing with 55-59. They computed accident rates in two ways: (1) per number of active pilots and (2) per number of annual hours flown. All analyses depend upon visually examining four figures displaying the above data; no statistical analyses were reported.

Three of the figures display the accidents per number of active pilots for air transport pilots, for commercial pilots, and for private pilots, respectively. Exposure was not accounted for in these figures. The fourth figure displays accidents per hours flown for Part 135 operators as a function of age and of hours flown in the preceding year (101 - 400 versus >400), although the data for the 20-24 year age group is omitted for some unstated reason. There appear to be no discernible trends as a function of age in the data in the fourth figure. Indeed, differences among the various age groups appear minimal, however, this is difficult to verify without statistical tests. For each of the seven age groups displayed, the pilots with more than 400 recent flight hours have an accident rate roughly one third that of the pilots with 101-400 recent flight hours. The probability of this happening by chance is less than 0.01 (applying the sign test, with $N=7$).

3.2.4 Replicating Golaszewski

Despite the limitations of Golaszewski's study, the general design has merit and deserves replication. The design could be replicated using data from the same period, 1976-1980, with design improvements incorporated based upon the criticisms described above so that reasonable interpretations of the data could be made relative to the Age 60 Rule. The design could be also be replicated using data from different periods, e.g., 1985-1989, and a determination made whether the same outcomes would occur with different age cohorts. Various ways to improve upon Golaszewski's work are discussed below.

Use statistical analyses. Foremost, the data should be analyzed statistically so that reasonable inferences can be made from the data. A number of statistical approaches are appropriate to this type of data analyses and the recommendations include: analysis of variance (ANOVA), regression, and chi square. Whereas the ANOVA and regression both assume normality, the chi square test is a non-parametric test which does not make any assumptions about the distribution of the date. As noted above, it is likely that accident rates are normally distributed and thus amenable to analysis using the first two approaches.

ANOVA and regression share the same underlying model and assumptions. Thus, either approach can be used to answer a given question, although some questions are more natural for a given approach, and the comments about ANOVA that follow apply to linear regression as well.

Initially, it appears that the ANOVA could be applied to a completely crossed, three factor design whose factors are age, recent flight time, and total flight time. However, in such a design only one observation per cell is available and the triple interaction of the factors could

35

not be tested. One way around this problem involves segregating the data by the third of a year (grouping months 1, 4, 7 and 10; months 2, 5, 8, and 11; and months 3, 6, 9, and 12) before computing accident rates. Then three observations per cell (year) are available, allowing the computation of an error term for the analysis. The ANOVA has the advantage of being able to answer specific questions about trends in the data. For example, the question of whether a specific age group has a different accident rate than the remaining groups could be posed.

The chi square analysis answers questions about the frequencies of occurrence in various categories. Accidents could be categorized and the chi square applied, however, the categories must be exhaustive. For example, to compare the number of flights with accidents as a function of age, the number of flights without accidents also must be tabulated and this data are not readily available. The chi square can be used to answer subsidiary questions about, for example, the distribution of the kind of accident as a function of age, class of medical certificate, etc.

Use smaller, more precise designs. Smaller designs that reach more definitive but limited conclusions are better than larger designs that reach more general but equivocal conclusions. Thus, grouping together all pilots and all accidents leads to only tentative conclusions. In the present situation, the analyses should bear more directly on the Age 60 Rule. The focus could be on Part 121 accidents only, which directly relates to the question, or on Class III pilots and Part 91 accidents involving Class III pilots to obtain a view of the general effects of age on accident rate. The data from Part 121 accidents may be too sparse for good statistical inferences. On the other hand, including the data from Class III pilots leads to analyses that are only indirectly relevant to the Age 60 Rule. In such a situation, it is appropriate to converge on the relevant question by using both approaches, with reasonable expectations of arriving at meaningful conclusions.

When examining data for other classes or flying types (e.g., Class III pilots and Part 91 accidents or Class II and Part 135 accidents), the data should be selected so that the population is reasonably comparable to pilots involved in Part 121 accidents. Golaszewski's data suggest a triple interaction of age, recent flight time, and total flight time. Thus, the analysis should restrict the population to pilots with high recent flight time (e.g., at least 200 hours) and high total flight time (e.g., at least 2000 hours), both of which are characteristic of professional airline pilots.

To get a clearer picture of the age effect, age should be categorized by finer intervals, e.g., one or two years, instead of the 10 year intervals used by Golaszewski. At the same time, to examine cohort effects, the data should be segregated by the year of the accident, rather than grouping together data for accidents occurring during some multiple-year interval.

For these data it is quite appropriate to use a completely crossed four-factor (age by recent-flight by total-flight by year-of-accident) ANOVA design. Planned comparisons for linear and quadratic trends can be used to test whether accident rate is a U-shaped function of age, an effect suggested by Golaszewski's study.

The results of the above analysis of Part 91 accidents are not fully generalizable to Part 121 accidents. The two populations of pilots differ in unknown ways. Nevertheless, the analysis of Part 91 accidents would provide information about the effects on aviation accidents

of aging in a less select population. Further, it would provide data for accidents by pilots older than 60 where the Part 121 accidents is unavailable. If accident rates increase at some point in time as one ages, the data from the analysis of Part 91 and 135 accidents should reflect this, especially when age is examined year by year.

Focusing on Part 121 accidents requires the identification of Part 121 accidents and requires good estimates of the exposure of pilots to these accidents. The NTSB Accident database directly identifies Part 121 accidents, but estimating exposure is more problematic. In addition to recording the number of hours flown, the FAA medical database records the pilot's employer, the pilot's job, and the kind of flying (solely business, solely pleasure, or both). Obtaining this information for all Class I pilots (the only pilots involved in Part 121 accidents), the number of hours of scheduled airline flights can be estimated.

Applying the ANOVA to Part 121 accidents exactly as recommended for Part 91 accidents suffers from the drawback that Part 121 accidents are rare, making the data sparse. This can be remedied by using 2-year intervals for age and by ignoring the year in which the accident occurred. Of course, this means that the data cannot be used to test the effect of cohort. Because the Age 60 Rule is predicated on changes occurring around 60 years, a planned comparison can be used to test the accident rate of the oldest age group (ages 59-60) against the remaining groups. This test both asks the right question and is statistically powerful.

3.2.5 Beyond Golaszewski

The above discussion is constrained by the kinds of data available in the present databases. The availability of additional information could provide more useful analysis modeled after that of Golaszewski. First, the data could be improved by using the same definition of "recent flight time" in the two databases. Because it is easier to recollect more recent events, the definition of "recent" as the last 90 days would be preferred. Second, pilots should separately report the amount of time engaged in the three categories (Part 121, Part 91, Part 135) of flying. This would avoid having to deal with the class of the medical certificate to estimate flight time and would yield much more precise estimates of accident rates. Third, the number of hours spent flying various makes and models of airplanes should be recorded in the medical database. These additional data would allow the analyses of the data that better discount factors not directly related to age. Finally, the number of landings should be recorded. This would allow investigators to study whether number of landings is a more appropriate variable than number of hours in computing accident rates.

Given the information in the two primary databases, what other kinds of analyses can be done? Forming a good estimate of exposure constitutes the toughest problem in the Golaszewski design. Analyses that avoid this problem cannot deal directly with the question of accident rate as a function of age, but they still illuminate the question. Based upon the information in the NTSB Accident database, accidents could be categorized in various ways. This would allow a determination of whether the frequency of accidents in the various categories differ as a function of age. The chi square would be quite appropriate for this analysis and the population of all accidents would be exhaustively partitioned into various categories. For example, whether the time of day of the accident varied as a function of age would be of interest. Then the chi square could be applied to a two factor ("time-of-day" by "age") frequency table of all accidents. In

a similar fashion, the analysis would focus on the information available in the CAIS medical database and ask questions about the trends in the various recorded medical variables. In some cases, ANOVA would be appropriate for the resulting data, in other cases the chi square would be appropriate.

The consolidated database (CDB) also provides the ability to answer questions about the integrity of the data. For example, the flight time recorded in the NTSB Accident database is based upon log entries, whereas the flight time recorded in the CAIS medical database is usually based upon the pilots' recollections. Are the latter data less reliable? Comparing the data from the two databases for the same pilots would answer this question.

3.2.6 Relationships among Medical Variables, Age, and Pilot Accidents

The CAIS medical database records the results of the periodic medical examinations required to maintain Class I, II, or III certification. The medical information in these records includes the following:

- Height and weight.

- Cardiovascular, including blood pressure, pulse at rest and after exercise, and results of electrocardiogram.

- Vision, including near and distant vision in each hemisphere, both corrected and uncorrected, color vision, and field of vision.

- Hearing, including for each ear testing by whisper and by audiometer at frequencies from 500 to 4000 hz.

- Results of test for albumin and sugar in urine.

- Physical abnormalities discovered during the examination.

- Medical pathologies from correspondence, from medical records, from remarks by applicant, or from AME's entries on application form.

- Answers to 14 Yes/No questions from medical history taken during the examination.

Brief summary of prior literature. The increased likelihood of sudden incapacitation with age partially constituted the original impetus for the Age 60 Rule (Lane, 1971). Consequently, there has been a great deal of research on those factors that lead to sudden incapacitation in pilots ("acute coronary events ..., new onset idiopathic epilepsy, and physiological problems including spatial disorientation, hypoxia,...") (Froom, Benbassat, Gorss, Ribak, & Lewis, 1988), even though a number of studies show that sudden incapacitation is very unlikely to cause an aviation accident (Booze, 1989; Martin-Saint-Laurent, Lavernhe, Casano, & Simkoff, 1990; Mohler & Booze, 1978; Rayman, 1973). Those studies that have considered the age of the pilot in conjunction with medical variables have mainly focused on cardiovascular

fitness. Bruce and Fisher (1987) suggest older pilots could be screened for cardiovascular risk. They readily identified men whose profile indicated either high or low risk for coronary disease. In a five year follow-up, 55-59 year olds with a high-risk profile had a mortality rate six times that of 60-64 year olds with a low-risk profile (34.2% versus 5.8%).

A number of authors have examined the effect of age on the pilot's vision or hearing. In these studies, with minor exceptions having to do with specific conditions, as pilot age increases, visual acuity decreases (Baisden & Monaco, 1983; Vola, Cornu, Carruel, Gastaud, & Leid, 1983), speed of accommodation increases (Elsworth, Larry & Malmstron, 1986; Temme & Morris, 1989), and the speed of shifting of gaze increases (Morris & Temme, 1989). The hearing losses in pilots is mostly a function of exposure to noise rather than of age (Fitzpatrick, 1988).

An annotated topical bibliography of prior literature linking key medical variables with age and pilot performance is presented in Appendix C. For most articles, the annotation consists primarily of a condensation of the abstract of the article.

3.2.7 Improved Ways to Examine Relationships among Medical Variables, Age, and Pilot Performance

The CDB provides medical information about all pilots, both those involved in accidents and those who are not. Thus, how well the medical variables predict an accident as a function of age, exposure, etc. can be tested. Such a prediction depends not only on knowing the characteristics of pilots involved in accidents but also on knowing the characteristics of pilots not involved in accidents.

The CDB records the results of tests of cardiovascular fitness, hearing, and vision. The predictive value of each of these variables could be tested as follows. First, determine the correlation of each measure with age. Then use logistic regression to determine how well the measure predicts accidents. For example, the correlation of cardiovascular fitness with age could be determined for this population. Then logistic regression can be used to determine how well age, exposure, and cardiovascular fitness predict an accident for each Class of pilot. This would shed light on the predictive value of the various measures. At the same time it would characterize some of the differences among the three classes of pilot.

3.3 Conclusions and Recommendations

Golaszewski made a worthwhile first step in analyzing the association of the age of the pilot with the probability of an accident. Given hindsight, numerous improvements can be made in Golaszewski's methodology. The Age 60 project will have the advantage of a CDB that combines the above two databases, providing measures of relationships unavailable to Golaszewski. Further, the CDB will provide the capability to generate more precise estimates of exposure (i.e., number of flight hours) and will provide checks on the reliability of pilots' reports of flight time. It is likely that number of landings is a better measure of exposure. This empirical question should be tested. If landings is, indeed, a better measure of exposure, then that data should be included in the information entered in the database.

Sudden incapacitation, while intensively studied, has been shown to be a relatively unimportant factor in aviation accidents. Thus, studies of the relationship between age and cardiovascular fitness in pilots are somewhat tangential to the question of whether the probability of an accident increases with age.

The data clearly show that vision of pilots deteriorates with age, while the hearing of pilots deteriorates with exposure to noise. Lacking a conceptual model of pilot performance, it is unclear what these data imply about the performance of pilots. For example, the older pilot with greater experience may more than compensate for weakened vision and hearing.

REFERENCES

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

Baisden, A. G., & Monaco, W. A. (1983). Distribution of visual characteristics of naval aviation personnel (Final Report). Pensacola, FL: Naval Aerospace Medical Research Lab.

Banich, M. T., Stokes, A. F., & Elledge, V. C. (1988). Neuropsychological screening of aviators: A review. Aviation, Space, and Environmental Medicine, 60, 361-366.

Banich, M. T., Stokes, A. F., & Karol, D. (1990). Cognitive function evaluation in the medical certification of airmen: Phase B (Final Technical Report ARL-90-2/FAA-90-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Barrett, G. V., Mihal, W. L., Panek, P. E., Sterns, H. L., & Alexander, R. A. (1977). Information processing skills predictive of accident involvement for younger and older commercial drivers. Industrial Gerontology, 4, 173-182.

Bartram, D. (1986). Development and evaluation of MICROPAT version 4. (TR-184). London: Ministry of Defense, and University of Hull, U.K.: Ergonomics Research Group.

Boer, L. C. (1986). Attention tasks and their relation to aging and flight experience (A83KLU078). Soesterberg, Netherlands: Institute for Perception.

Bohannon, R. L. (1969). Study of age 60 rule. Washington, DC: Federal Aviation Administration.

Booze, C. F. (1989). Sudden inflight incapacitation in general aviation. Aviation, Space, and Environmental Medicine, 60, 332-335.

Braune R., & Wickens C. D. (1984a). Individual differences and age-related performance assessment in aviators. Part 1: Battery development and assessment (Final Technical Report EPL-83-4/NAMRL-83-1). Urbana-Champaign, IL: University of Illinois, Engineering-Psychology Laboratory.

Braune, R., & Wickens, C.D. (1984b). Individual differences and age-related performance assessment in aviators Part 2: Initial battery validation (Final Technical Report EPL-83-7/NAMRL-83-2). Urbana-Champaign, IL: University of Illinois, Engineering Psychology Laboratory.

Braune, R., & Wickens, C.D. (1985). The functional age profile: An objective decision criterion for the assessment of pilot performance capacities and capabilities. Human Factors, 27, 681-693.

Bruce, R. A., & Fisher, L. D. (1987). Exercise-enhanced risk factors for coronary heart disease vs. age as criteria for mandatory retirement of healthy pilots. Aviation, Space, and Environment Medicine, 58, 792-798.

Carretta, T. R. (1987a). Basic attributes tests (BAT) system: Development of an automated test battery for pilot selection (AFHRL-TR-87-9). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cerella, J. (1990). Aging and information-processing rate. In J. E. Birren, & K. W. Schaie (Eds.), Handbook of the psychology of aging (pp. 201-219). San Diego: Academic Press.

Cerrella, J., Poon, L. W., & Williams, D. (1980). Age and the complexity hypothesis. In L. W. Poon (Ed.), Aging in the 1980's (pp. 332-340). Washington, DC: American Psychological Association.

Chappell, S. L. (1991). Training and cockpit design to promote expert performance. In R. S. Jensen (Ed.), Proceedings of the sixth international symposium on aviation psychology (pp. 133-138). Columbus, OH: Ohio State University, Department of Aviation, Aviation Psychology Laboratory.

Charness, N. (1983). Age, skill, and bridge bidding: A chronometric analysis. Journal of Verbal Learning and Verbal Behavior, 22, 406-416.

Charness, N. (1985). Introduction to aging and human performance. In N. Charness (Ed.), Aging and human performance. New York: Wiley.

Charness, N. (1989). Age and expertise: Responding to Talland's challenge. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), Everyday cognition in adulthood and late life. (pp. 437-456). Cambridge: Cambridge University Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

Chidester, T. R. (1990). Human factors research: Narrowing the extremes of flight crew performance. Washington, DC.: Federation of Behavioral, Psychological and Cognitive Sciences: Science and Public Policy Seminars.

Childs, J. M., & Spears, W. D. (1986). Flight-skill decay and recurrent training. Perceptual and Motor Skills, 62, 235-242.

Collins, W. E., & Mertens, H. W. (1988). Age, alcohol, and simulated altitude: Effects on performance and breathalyzer scores. Aviation, Space, and Environmental Medicine. 59, 1026-1033.

Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren &
    K. W. Schaie (Eds.), Handbook of the Psychology of Aging (pp. 384-420). New York:
    Van Nostrand Reinhold.

Damos, D., & Gibb, G. (1986). Development of computer-based naval aviation
    selection test battery (NAMRL-1319). Pensacola, FL: Naval Aerospace Medical
    Research Laboratory.

DeGroot, A. (1965). Thought and choice in chess. The Hague: Mouton.

Drake, A. W. (1967). Fundamentals of applied probability. New York: McGraw Hill.

Elworth, C. L., Larry, C., & Malmstrom, F. V. (1986). Age, degraded viewing environments,
    and the speed of accommodation. Aviation, Space, and Environmental Medicine, 57,
    54-58.

Endsley, M. R. (1989). A methodology for the objective measurement of
    situation awareness. Paper presented at the AGARD Symposium on Situation Awareness
    in Aerospace Operations, Copenhagen.

Federal Aviation Administration (1986). Airplane simulator and visual system
    evaluation (FAA/DOT/AC-120-40A). Washington, DC.

Federal Aviation Administration (1990). Line operational simulations: Line-
    oriented flight training, special purpose operational training line operational evaluation
    (DOT/FAA/AC-120-35B). Washington, DC.

Federal Aviation Administration (1991). Aeromedical certification statistical
    handbook. Washington, DC.

Fitzpatrick, D. T. (1988). An analysis of noise-induced hearing loss in Army
    helicopter pilots. Aviation, Space, and Environmental Medicine, 59, 937-941.

Fozard, J. L., Vercruyssen, M., Reynolds, S. L., & Hancock, P. A. (1990).
    Longitudinal analysis of age-related slowing: BLSA reaction time data. Proceedings of
    the Human Factors Society 34th Annual Meeting (pp. 163-167).

Froom, P., Benbassat, J., Gross, M., Ribak, J., & Lewis, B. S. (1988). Air
    accidents, pilot experience, and disease-related inflight sudden incapacitation. Aviation,
    Space, and Environmental Medicine. 59, 278-281.

Gerathewohl, S. J., (1977). Psychophysiological effects of aging: Developing a
    functional age index for pilots: I. A survey of the pertinent literature (Report No. FAA-
    AM-77-6). Washington, DC: FAA Office of Aviation Medicine.

Gerathewohl, S. J. (1978a). Psychophysiological effects of aging: Developing a functional age index for pilots: II. Taxonomy of psychological factors (OAM Report No. FAA-AM-78-16). Washington, DC: FAA Office of Aviation Medicine. (NTIS No. 3985748).

Gerathewohl, S. J. (1978b). Psychophysiological effects of aging: Developing a functional age index for pilots: III. Measurement of pilot performance (Report No. FAA-AM-78-27). Washington, DC: FAA Office of Aviation Medicine. (NTIS No. HC A04/MF A01)

Glanzer, M., & Glaser, R. (1959). Cross-sectional and longitudinal results in a study of age-related changes. Educational and Psychological Measurement, 19.

Golaszewski, R. (1983). The influence of total flight time, recent flight time and age on pilot accident rates. Washington, DC: FAA Office of Aviation Safety.

Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. Perceptual and Motor Skills, 33, 1335-1342.

Green, R. G. (1985). Stress and accidents. Aviation, Space, and Environmental Medicine, 56, 638-641.

Gregorich, S. E. (1991). What makes a good loft scenario?: Issues in advancing current knowledge of scenario design. In R. S. Jensen (Ed.), Proceedings of the Sixth International Symposium on Aviation Psychology (pp. 981-985), Columbus, OH: Ohio State University, Department of Aviation, Aviation Psychology Laboratory.

Guide, P. C. & Gibson, R. S. (1991). An analytical study of the effects of age and experience on flight safety. Proceedings of the Human Factors Society 35th Annual Meeting (pp. 180-184).

Guilford J. P. (1947). Printed classification tests (Army Air Forces Aviation Psychology Research Report No. 5). Washington, DC: U.S. Government Printing Office.

Guilliom, D. C., Spears, W. D., Demuth, J. J., Eddy, P. O., Hanley, D. E., Holmbeck, G. E., & Fishburne, R. P. (1985). A systematic determination of skill and simulator requirements for airline transport pilot certification (Final Report DOT-TSC-FAA-85-1). Falls Church, VA: Planning Systems International, Inc.

Halpern, D. F. (1984). Age differences in response time to verbal and symbolic traffic signs. Experimental Aging Research, 10, 201-204.

Helmreich, R. L., Savin, 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, 185-188.

Horst, R. L., & Kay, G. G. (1991). COGSCREEN: Personal computer-based tests of cognitive junction for occupational medical certification. In R. S. Jensen (Ed.), Proceedings of the Sixth International Symposium on Aviation Psychology (pp. 733-739). Columbus, OH: Ohio State University, Department of Aviation, Aviation Psychology Laboratory.

Hoyer, W. (1987). Domains of attention. Paper presented at the National Institute on Aging Conference on Aging and Attention, Washington, DC.

Hultsch, D. F. & Dixon, R. A. (1990). Learning and memory in aging. In J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging (pp. 259-269) San Diego: Academic Press.

Hunt, E., & Hertzog, C. (1981). Age related changes in cognition during the working years (Final Report). Arlington, VA: Personnel and Training Research Programs, Office of Naval Research.

Imhoff, D. L., & Levine, J. M. (1981). Perceptual motor and cognitive performance task battery for pilot selection (AFHRL-TR-75-10). TX: Brooks Air Force Base.

Institute of Medicine, National Academy of Sciences (1981). Airline pilot age, health and performance. Washington, DC: National Academy Press.

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

Jensen, R. S., & Benel, R. A. (1977). Judgement evaluation and instruction in civil pilot training (Final Report FAA-RD-78-24). Springfield, VA: National Technical Information.

Klein, G. A. (1989). Recognition-primed decisions. In W.R. Rouse (Ed.), Advances in man-machine systems research, Vol. 5.

Kline, D. W., & Birren, J. E. (1975). Age differences in backward dichotic masking. Experimental Aging Research, 1, 17-25.

Kline, D. W., & Szafran, J. (1975). Age differences in backward monoptic visual noise masking. Journal of Gerontology, 30, 307-311.

Kruk, R., Regan, D., Beverley, K. I., & Longridge, T. (1978). Flying performance on the advanced simulator for pilot training and laboratory tests of vision. Human Factors, 25(4), 456-457.

Lane, J. C. (1971). Risk of in-flight incapacitation of airline pilots. Aerospace Medicine, 42, 1319-1321.

Leirer, V. O., Yesavage, J. A., & Morrow, D. G. (1989). Marijuana, aging, and task difficulty effects on pilot performance. Aviation, Space, and Environmental Medicine, 60, 1145-1152.

MacIntyre, N. R., Mitchell, R. E., Oberman, A., Harlan, W. R., Graybiel, A., & Johnson, E. (1978). Longevity in military pilots: 37-year follow-up of the Navy's "1000 aviators". Aviation, Space, and Environmental Medicine, 49, 1120-1122.

Martin-Saint-Laurent, A., Lavernhe, J., Casano, G., & Simkoff, A. (1990). Clinical aspects of inflight incapacitations in commercial aviation. Aviation, Space, and Environmental Medicine, 61, 256-260.

McDowd, J. M., & Birren, J. E. (1990). Aging and attentional processes. In J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging (pp. 222-230). San Diego: Academic Press.

McGhie, A. (1969). Pathology of attention. London: Penguin.

Mertens, H. W., Higgins, E. A., & McKenzie, J. M. (1983). Age, altitude, and workload effects on complex performance. Oklahoma City, OK: Civil Aeromedical Institute, FAA. (NTIS No. FAA-AM-83-15).

Modell, J. G., & Mountz, J. M. (1990). Drinking and flying: The problem of alcohol use by pilots. New England Journal of Medicine, 323, 455-461.

Mohler, S. R. (1981). Reasons for eliminating the "Age 60" regulation for airline pilots. Aviation, Space, and Environmental Medicine, 52, 445-454.

Mohler, S. R., & Booze, C. F. (1978). U.S. fatal general aviation accidents due to cardiovascular incapacitation: 1974-75. Aviation, Space, and Environmental Medicine, 48, 944-948.

Morris, A., & Temme, L. A. (1989). Time required for U.S. Navy fighter pilots to shift gaze and identify near and far targets. Aviation, Space, and Environmental Medicine, 60, 1085-1089.

Morrow, D., Leirer, V., & Yesavage, J. (1990). The influence of alcohol and aging on radio communication during flight. Aviation, Space, and Environmental Medicine, 61, 12-20.

Mosier, K. L. (1991). Expert decision-making strategies. In R. S. Jensen (Ed.), Proceedings of the Sixth International Symposium on Aviation Psychology (pp. 266-271). Columbus, OH: Ohio State University, Department of Aviation, Aviation Psychology Laboratory.

National Institute on Aging (1981). Report of the National Institute on Aging panel on the experienced pilots study. Bethesda, MD: NIH, NIA.

North, R. A., & Gopher, D. (1976). Measures of attention as predictors of flight performance. Human Factors, 18, 1-14.

Office of Technology Assessment (1990). Medical risk assessment and the age 60 rule for airline pilots. Washington, DC: Subcommittee on Investigations and Oversight, Committee on Public Works and Transportation, U.S. House of Representatives.

Parasuraman, R., & Giambra, L. (1991). Skill development in vigilance: Effects of event rate and age. Psychology and Aging, 6, 155-169.

Perez, W. A., Masculine, P. L., Ramsey, E. G., & Urban, K. E. (1987). Unified tri-services cognitive performance assessment battery: Review and methodology (AAMRL-TR-87-007). Wright-Patterson, AFB, OH: Armstrong Aeromedical Research Laboratory.

Plude, D. J., & Hoyer, W. J. (1985). Attention and performance: Identifying and localizing age defects. In N. Charness (Ed.), Aging and human performance (pp. 97-99). New York: Wiley.

Ponds, R. W. H. M., Brouwer, W. H., & van Wolffelaar, P. C. (1988). Age differences in divided attention in a simulated driving task. Journal of Gerontology: Psychological Sciences, 43, 151-156.

Proper, R. (1969). Study of physiological and psychological aging in pilots progress report (National Institute of Health, Grant No. HD-00518). Albuquerque, NM: The Lovelace Foundation for Medical Education and Research.

Ross, L. E., & Mundt, J. C. (1988). Multiattribute modeling analysis of the effects of a low blood alcohol level on pilot performance. Human Factors, 30, 293-304.

Rowe, J. W., & Kahn, R. L. (1987). Human aging: Usual and successful. Science, 237, 143-149.

Rybash, J. M., Hoyer, W. J., & Roodin, P. A. (1987). Adult cognition and aging: Developmental changes in processing, knowing and thinking. New York: Pergamon Press.

Salthouse, T. A. (1985). A theory of cognitive aging. Amsterdam: North Holland.

Salthouse, T. A. (1987). Age, experience, and compensation. In C. Schooler & K. W. Schaie (Eds.), Cognitive functioning and social structure over the life course (pp. 142-157). Norwood, NJ: Ablex.

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by
        reductions in working memory and speed of processing. Psychological Science, 2, 179-
        184.

Salthouse, T. A., & Somberg, B. L. (1982). Isolating the age deficit of
        speeded performance. Journal of Gerontology, 37, 59-63.

Samet, M., Geiselman, R. E., Zajaczkowski, F., & Marshall-Miles, J. (1986).
        Complex cognitive assessment battery CCAB): Test descriptions. Alexandria, VA: U.
        S. Army Research Institute.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined
        phenomenon. The International Journal of Aviation Psychology, 1, 45-57.

Schaie, K. W. (1973). Methodological problems in descriptive developmental
        research on adulthood and aging. In J. R. Nesselroade & H. W. Reese (Eds.), Life-span
        developmental psychology: Developmental issues. New York: Academic Press.

Schaie, K. W. (1977). Quasi-experimental research designs in the psychology
        of aging. In J. E. Birren and K. W. Schaie (Eds.), Handbook of the
        psychology of aging. New York: Van Nostrand Reinhold.

Schaie, K. W. (1986). Beyond calendar definitions of age, period, and cohort:
        The general developmental model revisited. Developmental Review, 6, 252-277.

Schonfield, D. H. (1981). Attention switching in higher mental processes. In
        F. I. M. Craik & S. Trehub (Eds.), Aging and cognitive processes. New York: Plenum.

Scialfa, C. T., Guzy, L. T., Leibowitz, H. W., Garvey, P. M., & Tyrrell, R. A.
        (1991). Age differences in estimating vehicle velocity. Psychology and Aging, 6, 60-66.

Stokes, A. F. (1991). MIDIS - A microcomputer flight decision simulator. In
        E. Farmer (Ed.), Human Resource Management in Aviation. Alderslot: Avebury
        Technical.

Stokes A. F., Banich M. T., & Elledge, V. (1988). Cognitive function evaluation in
        the medical certification of airmen (Final Tech. Rep. ARL-88-4/FAA-88-2). Savoy, IL:
        University of Illinois, Aviation Research Laboratory.

Stokes, A. F., Belger, A. & Zhang, K. (1990). Investigation of factors
        comprising a model of pilot decision making: Part II. Anxiety and cognitive strategies
        in expert and novice aviators (Final Technical Report ARL-90-8/SCEEE-90-2). Wright-
        Patterson AFB, OH: AMRL, Human Engineering Division.

Stokes, A. F., & Raby, M. (1989). Stress and cognitive performance in trainee pilots. Proceedings of the 33rd Annual Meeting of the Human Factors Society (pp. 883-887). Santa Monica, CA: Human Factors Society.

Stokes, A., Wickens, C., Davis, Jr., Barnett, B., Rosenblum, R., Hyman, F. (1991). A study of pilot decision making using MIDIS - a microcomputer-based flight decision training system. In R. S. Jensen (Ed.), Proceedings of the Sixth International Symposium on Aviation Psychology. Columbus, OH: Ohio State University, Department of Aviation, Aviation Psychology Laboratory.

Stones, M. J., & Kozma, A. (1989). Physical activity, age and cognitive/motor performance. In M. L. Howe & C. J. Brainerd (Eds.), Cognitive development in adulthood: Progress in cognitive development research. New York: Springer.

Szafran, J. (1966). Age differences in the rate of gain of information, signal detection strategy and cardiovascular status among pilots. Gerontologia, 12, 6-17.

Szafran, J. (1969). Psychological studies of aging in pilots. Aerospace Medicine, 40, 543-553.

Temme, L. A., & Morris, A. (1989). Speed of accommodative and age. Optometry and Vision Science, 66, 106-112.

Temple, L. L. (1989). Perceptual and cognitive factors affecting driving ability of young and old drivers. Las Vegas, NV: University of Nevada, Department of Psychology.

Tsang, P. S. (1990). A reappraisal of aging and pilot performance. Proceedings of the Fifth International Symposium of Aviation Psychology. Dayton, OH.

Ursano, R. J. (1980). Stress and adaptation: The interaction of the pilot personality and disease. Aviation, Space, and Environmental Medicine. 51, 1245-1249.

Vola, J. L., Cornu, L., Carruel, C., Gastaud, P., & Leid, J. (1983). Age and photopic and mesopic visual acuity. Journal of French Ophthalmology, 6, 473-479.

Waag, W. L. (1988). Programs and prospects in aircrew performance measurement. Aviation, Space, and Environmental Medicine, 59(11), suppl.), A46-A51.

Welford, A. T. (1964). Experimental psychology in the study of aging. British Medical Bulletin, 20, 65-69.

Wickens, C. D., Braune, R., & Stokes, A. (1987). Age differences in the speed and capacity of information processing: 1. A dual-task approach. Psychology and Aging, 2, 70-78.

Wickens, C. D., Braune, R., Stokes, A., & Strayer, D. (1985). Individual differences and age related change: Refinement and elaboration of an information processing performance battery with aviation relevant task structure (Final Report EPL-85-1/NAMRL-85-1). Champaign, IL: University of Illinois Engineering Psychology Laboratory, Department of Psychology.

Wickens, C. D., & Flach, J. (1988). Information processing. In E. Wiener & D. Nagel (Eds.), Human Factors in Aviation. New York: Academic Press.

Wickens, C. D., Stokes, A., Barnett, A. F., & Davis, T. (1988). A componential analysis of pilot decision making. Wright-Patterson AFB, OH: Technical Report AAMRL-TR-88-017.

Yesavage, J. A., Leirer, V. O., Denari, M., & Hollister, L. E. (1985). Carry-over effects on marijuana intoxication on aircraft pilot performance: A preliminary report. American Journal of Psychiatry, 142, 1325-1329.

APPENDIX A

Test Batteries Used for Pilot Selection or Evaluation

Ernest B. Gurman
University of Southern Mississippi

CONTENTS
APPENDIX A

Contents
Page 2

COGSCREEN

PURPOSE

The COGSCREEN test battery was designed to be used as a cognitive function screening test for medical certification of commercial pilots. The purpose was to provide an efficient computer-based means to evaluate the components of cognitive functioning believed to be important to aviation safety. It was intended that COGSCREEN would reveal cognitive deficits in aviators more readily than would observation of actual flight under normal conditions. The COGSCREEN tests were not intended to be a comprehensive neuropsychological evaluation battery.

DESCRIPTION

The COGSCREEN test battery was developed over a period of several years by Horst and Kay (1991). The sub-tests were selected, based on task analyses available in the human factors literature. The COGSCREEN tests do not require knowledge of aviation tasks or familiarity with computer technology. The current version of COGSCREEN consists of a battery of thirteen cognitive tests that tap the mental processes involved in working memory, associative memory, selective attention, time sharing, and visual-spatial, verbal-sequential, and psychomotor functioning.

The battery was designed to be used repeatedly with a given subject. Alternate forms of the sub-tests can be computer generated to minimize practice effects. The COGSCREEN sub-tests are as follows.

Backward Digit Span

Groups of three to six digits are presented sequentially. Subjects are required to reproduce each sequence in reverse order. The performance score reflects the longest digit span that the subject successfully reproduces at least one sequence and the total number of sequences correct. This test taps attention, working memory, and verbal-sequential processing.

Visual Sequence Comparison

Pairs of alphanumeric strings, four to eight characters in length, are presented simultaneously on the monitor. The subject indicates "same" or "different" for each pair of strings. "Same" means the same characters in the same positions. Performance is measured by both response speed and accuracy. This test measures attention, short-term visual memory, and verbal-sequential processing.

Auditory Sequence Comparison

Pairs of four to eight tones, varying in pitch, are presented sequentially. The subject indicates "same" or "different" for each pair of sequences. Performance is measured by both

response speed and accuracy. This test taps attention, short-term auditory memory, and sound pattern sequencing.

Math Problems

These are word problems with multiple choice answers that measure simple mental arithmetic. Performance is measured by both response speed and accuracy. This test taps working memory, long-term memory, mental arithmetic, and logical reasoning.

Manikin Figures

A picture of a man holding a flag in one hand is presented on the monitor in one of four orientations (i.e., right side up or upside down, and facing inward or outward from the screen). The subject must indicate which hand the manikin is holding the flag. Performance is measure by both response speed and accuracy. This test taps visual-spatial orientation and the ability to rotate mental images.

Symbol Digit Coding

Six paired symbols and digits are displayed at the top of the monitor throughout the test. Below this display a row of symbols are presented in random order, with adjacent blank spaces. The subject fills in the appropriate digit for each symbol by referring to the display at the top of the screen. Performance is measured by the number of items completed in 90 seconds and the percentage correct. This test taps attention, visual scanning, working memory, and verbal-sequential processing.

Symbol Digit Coding - Immediate Recall, Delayed Recall

Immediately after the Symbol Digit Coding Test, and again after a 30 minute delay, the six symbols are presented in random order and the subject must recall the digits that had been paired with each symbol. Performance is measured by the number of correct items. This test taps immediate and delayed recall.

Matching to Sample

A sample grid pattern with filled and empty cells is presented briefly. After a short delay the same pattern is presented with a slightly different pattern. The subject indicates the grid pattern that is the same as the previously presented pattern by a forced-choice response. Performance is measured by both response speed and accuracy. This test taps visual memory and visual-spatial processing.

Divided Attention

In the upper half of the screen a horizontal bar moves up and down within a circle, changing direction at unpredictable times. The subject is instructed to respond when the moving bar passes from the center of the circle into the upper or lower colored regions. A response returns the bar to the center of the circle. This task is performed alone and concurrently with

a Visual Sequence Comparison task. Performance is measured by the speed and accuracy of responses to each task, and the differences between single task and dual task conditions. In the dual task mode, this test taps divided attention, working memory, visual-spatial processing, and verbal-sequential processing.

Pathfinder

A character is displayed in each the four quadrants of the screen. From one display to the next these characters form a sequence. With each new display, the subject indicates the character that comes next in sequence. The task is performed with the sequence of numbers, letters, and an alternating sequence of numbers and letters. Performance is measured by the time to complete each sequence and the spatial accuracy of responses. This test taps verbal-sequential processing, working memory, attention, and the ability to apply rules systematically.

Shifting Attention

Four response boxes are displayed near the bottom of the screen, one with a colored border, one with an uncolored border, one with no border but containing a colored arrow pointing right, and one with no border but with an uncolored arrow pointing left. Stimuli are displayed in a similar box above this row of response boxes. There are five conditions, each requiring the subject to respond according to different rules. (1) The Border Color condition requires a straightforward matching of the stimulus to response boxes, whereas the (2) Arrow Direction and (3) Arrow Color conditions introduce the interference of an irrelevant stimulus dimension. In the (4) Instruction Condition, the response rule changes from trial to trail, with a message informing the subject of the rule for the upcoming trial. In the (5) Discovery condition the rule that is in effect must be discovered by trial and error. After an unpredictable number of consecutive correct responses, the rule is changed. In all conditions, the performance is measured by both response speed and accuracy. This test taps attention, visual discrimination, concept formation, and the systematic application of rules.

Dual task Tracking

There are two tasks, each is performed separately and then concurrently. One is a second-order compensatory tracking task in which the subject depresses the right and left arrow keys on the keyboard to center a vertical bar moving along a horizontal line. In the other task, three numbers are presented sequentially in random order. The subject indicates with the light pen which number was presented previously. Performance is measured by the average absolute error and number of boundary hits in the Tracking task and the response speed and accuracy in the Previous Number task. Together, these two tasks tap fine motor speed and coordination, visual-spatial and verbal-sequential processing, working memory, and divided attention.

PUBLISHER/AGENCY

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma

ADMINISTRATION

Time

Approximately 30-45 minutes are required to complete the test.

Apparatus required

COGSCREEN runs on a standard MS-DOS IBM PC-AT compatible computer with color monitor. The battery uses a light pen as the primary input device.

Procedure

The COGSCREEN software was designed to support both standardized test administration and special-purpose research applications. Individual tests can be called from a menu or the entire battery can be run in an automated stream. In the automated mode, the battery is self-administered. Each test includes on-line instructions and a brief practice session in which the subject must demonstrate sufficient proficiency to indicate that the instructions were understood. Trial data and some derived performance measures are saved to disk. The selected tests all lend themselves to alternate forms. The session number entered by the test administrator at run-time serves as the random number seed for automatically generating stimuli. The battery can be administered repeatedly to the same individuals with little test-retest bias.

Since COGSCREEN was developed for use with pilots as the target population, the test instructions appear on the screen with no provisions for any mental or physical limitations. It was assumed that the test population would be well-educated, highly motivated, and would have good visual acuity and sensory-motor ability.

The light pen input device was chosen to minimize biases in favor of subjects who may be familiar with computer keyboards. The use of a pointing device and labelled response boxes on the screen was intended to aid the explanation of the test instructions. The battery was designed to be less intimidating by allowing the subject to review the instructions and offering trial-by-trial feedback during the practice segments.

Reliability

Reliability studies are not available. Practice effects are being examined currently.

Criteria

Not available

Validity

Validity studies are not available. Age related biases are being examined currently.

Normative Data

COGSCREEN is currently being standardized on commercial pilots.

Scoring

Performance data is automatically logged and scored. A supplemental program can compute additional measures and create an output file that can be ported into a statistical analysis software package.

Summary

The COGSCREEN battery is under development and reliability and validity studies are not available. However, similar computer-based screening devices have been examined e. g., Stokes, Banich, and Elledge, (1991). They concluded "mini-mental exams" are poor candidates for large scale screening of line pilots for cognitive impairment. This conclusion was in part based on the low prevalence of cognitive deficits among commercial pilots.

The research results that have been reviewed on the use of computer-based tests for identifying cognitive deficits show that the tests may be useful in predicting membership in a population (e. g., differentiating pilots from known mentally impaired persons). However, the tests have not been demonstrated to be useful for differentiating among normally functioning pilots. The tests need further refinement to be able to make the accurate discriminations that are required.

- Advantages

    - COGSCREEN takes approximately 30-45 minutes to administer.

    - The battery can be self-administered.

    - The battery does not require special hardware and can be run on readily available microcomputers.

    - Alternate forms of the sub-tests can be computer generated.

    - Performance data is automatically logged and scored.

- Disadvantages

    - The battery has yet to be validated for detecting cognitive impairment in experienced pilots.

REFERENCES

Horst, R. L., & Kay, G. G. (1991). COGSCREEN: Personal computer-based tests of cognitive function for occupational medical certification. In R. S. Jensen (Ed.), Proceedings of the Sixth International Symposium on Aviation Psychology. Columbus, Ohio.

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.

BASIC ATTRIBUTES TEST

PURPOSE

The Basic Attributes Test (BAT) is a microcomputer-administered battery of tests developed to assess a candidates's potential for successfully completion of U.S. Air Force basic flying training. The BAT battery was designed to measure psychomotor skills, cognitive and perceptual abilities, personality characteristics, and attitudes believed to be related to flying training performance. The BAT battery was designed to measure a variety of psychomotor skills, information processing abilities, and personality characteristics that are considered important in determining the suitability of a candidate for flight training.

DESCRIPTION

Each test in the BAT battery was adapted from tests in the research literature that were identified as potentially useful predictors of flying performance (Imhoff & Levine, 1981). The criteria used to select the tests included feasibility, interest of the test-taker, independence from other tests in the battery, construct validity, and minimal dependence on verbal materials for administration. Individual test summaries are presented below. BAT apparatus specifications and details regarding hardware and software development are provided in Carretta (1987a).

Test Battery Introduction

This interactive subprogram prompts the subject to provide background information (e.g., identity, age, gender), also personal history and attitudes related to flying.

Two Hand Coordination

Two tests are used to evaluate psychomotor abilities. The first, the Two-Hand Coordination Test, is a rotary pursuit task. An airplane (target) moves in an elliptical path on the screen at a rate of 20 cycles per minute. The rate of movement of the airplane within each cycle varies in a fixed sinusoidal pattern. The subject controls the movement of a small "gunsight" using a left-hand joystick for vertical movement of the gunsight and a right-hand joystick for horizontal movement of the gunsight. The subject's task is to keep the gunsight on the moving airplane. After receiving instructions, the subject completes a 3-minute practice session and a 5-minute test. The measures of interest are horizontal and vertical tracking error scores. The psychological factors assessed in the Two-Hand Coordination Test are low-to-moderate-order tracking and time-sharing ability in pursuit.

Complex Coordination

The Complex Coordination Test uses a dual-axis joystick (right-hand joystick) to control the horizontal and vertical movement of a cursor. The left-hand single-axis joystick controls the horizontal movement of a "rudder bar" at the base of the screen. The subject's task is to keep the cursor (against a constant horizontal and vertical rate bias) centered on a large cross fixed at the center of the screen, while simultaneously centering the rudder bar at the base of the

screen (also against a constant rate bias). The instructions, practice, testing, and scoring are the same as those in the Two-Hand Coordination Test. The complex Coordination Test assesses compensatory tracking ability involving multiple-axis continuous events.

Encoding Speed

Two letters are presented simultaneously to the subject, who is required to make a same-different judgment about the letter pair. The judgment may be based on a Physical identity rule (look the same or look different, AA versus Aa), a Name identity rule (same name or different name, AA versus AH) or Category identity rule (vowels versus consonants - AE versus AH). The reaction time for the judgment provides a measure of the speed of the cognitive encoding process. Reaction time and accuracy of response (correct/incorrect) are recorded for each trial (32 trials in each rule condition for a total of 96 trials). The psychological factor involved in this test is verbal classification at several levels of cognitive operation.

Mental Rotation

A pair of letters is presented sequentially, and the subject is required to make a same-different judgment. Elements of the letter pair may be either identical or mirror images, and the letters may be either in the same orientation or rotated in space with respect to each other. A correct "different" judgment occurs when one element of a pair is a mirror image of the other, regardless of orientation.

To take the test, the subject must form a mental image of the first letter (no longer displayed) and perform a point-by-point comparison with the second letter (which remains on the screen). In addition, when the letters are rotated with respect to each other, the subject must mentally rotate the mental image of one letter into congruence with the other before making the comparison. Reaction time and accuracy of response are recorded on all 72 trials. The psychological factors assessed by this test are spatial transformation and classification.

Item Recognition

In this test, a string of one to six digit is presented on the screen. The string is then removed and followed, after a brief delay, by a single digit. The subject is instructed to remember the initial string of digits, then to decide if the single digit was one of those presented in the initial string. The subject is instructed to press a keypad button marked YES if the single digit was in the string, or press another marked NO if the digit was not in the string. The instructions inform the subject to work as quickly and accurately as possible. Reaction time and accuracy of response are recorded on all 48 trials. Short-term memory storage, search, and comparison operations are the underlying psychological factors for this test.

Time-Sharing

During a series of 10 1-minute trials, the subject is required to learn a compensatory tracking task. To perform this task, the subject must anticipate the movement of a gunsight on a screen and operate a control stick to counteract that movement to keep the gunsight aligned with a fixed central point (an airplane). Task difficulty is adjusted throughout the test,

depending on the subject's performance. The gunsight movement control dynamics are a combination of rate and acceleration components. The "disturbance" factor is a quasi-random summed sinusoidal forcing function.

After these "tracking only" trials, the subject is required to track while canceling digits that appear at random intervals and locations on the screen. A digit is canceled when the subject presses the corresponding button on the keypad. If the subject fails to respond to a digit within 4 seconds after its appearance time, the gunsight will disappear until a digit response is made. These dual-task trials occur in two 3-minute blocks. The information processing load gradually increases during these trials. The Time-Sharing Test ends with a final 3-minute block of "tracking only" trials. There are a total of 19 1-minute trials(10 tracking only, 6 dual-task, and 3 more tracking only). The effect of the secondary task loads is reflected in the pattern of level of tracking difficulty changes caused by the adaptive logic that holds tracking error constant. Feedback concerning tracking difficulty is provided by a gauge that appears in the top right of the screen. The measures of interest for this test include the level of tracking difficulty at which the subject can perform consistently, response time on the secondary task, and dual-task performance. This test assesses a variety of psychological factors including higher-order tracking ability, and learning rate and time-sharing ability as a function of differential task load (Carretta, 1987b).

Self-crediting Word Knowledge

This test is essentially a vocabulary test in which a "target" word is presented to the subject along with five other words from which its closest synonym must be chosen. There are three blocks of 10 questions each. The target words become more difficult with each successive block. Subjects are informed of this increasing difficulty and are required to make a "bet" before each block according to how they expect to do. Response time and accuracy are recorded on all 30 trials. This test assesses self-assessment ability and self-confidence.

Activities Interest Inventory

This test is designed to determine the subject's interest in a variety of activities. Eighty one pairs of activities are presented, and the subject is asked to choose between them. Subjects are told to assume they have the necessary ability to perform each activity. The activity pairs force the subjects to choose between tasks that differ as to degree of threat to physical survival (sometimes subtly, sometimes not).The measures of interest are the number of high-risk options chosen and the amount of time required to choose between pairs of activities. The psychological factor assessed by this test is attitude toward risk-taking.

PUBLISHER/AGENCY

Human Resources Directorate Manpower and Personnel Division.
Air Force Systems Command
Brooks AFB, TX 78235-5000

RELIABILITY

Test-retest reliability with 247 USAF pilot candidates on two consecutive days yielded a Pearson r = .56. It was concluded that short-term retesting on the BAT should be avoided to minimize test score gains due to prior exposure to the test, (Carretta, 1991).Internal consistency reliability estimates were also computed. Chronbach's alpha ranged between .95 to .41 for the various measures. These correlations were on a shortened nine subtest version of the BAT. The abbreviate version was used because of time constraints and required about 2 hours to complete. Subtest reliability results are presented in Table 1.

CRITERIA

The criterion used in the validity study was Undergraduate Pilot Training (UPT) pass/fail final outcome.

VALIDITY

The validity of the BAT was studied and reported in combination with the Air Force Officers Qualifying Test (AFOQT). The subjects were 455 United States Air Force (USAF) UPT students. The best combination of predictors was achieved through a simultaneous regression solution $r = .227$. $p < .01$, N =430 and $r = .242$, $p < .01$, N = 455, (Carretta, 1990b).

ADMINISTRATION

Time

The BAT requires about 3 hours to complete. The test session is self-paced by the subject. Rest periods are programmed between tests.

Apparatus

The BAT is based on a microcomputer and monitor built into a rugged chassis with a glare shield and side panels designed to minimize distractions.

Procedure

Each subject responds to the tests by using individually or in combination a two-axis joystick on the right side of the apparatus, a single-axis joystick on the left side, and a keypad in the center of the test unit. The keypad includes the numbers 0 to 9, an "ENABLE" key in the center, and a bottom row with "YES" and "NO" keys and two others labeled "S/L" (for same/left responses) and "D/R" (for different/right responses).

Norms

Norms would need to be developed for the target population.

Scoring

Scoring procedures for the BAT battery rely on a combination of tracking error and difficulty, response speed, response accuracy and response choice, and sometimes, specially derived scores (e.g.,regression slope and intercept, interaction terms).

TWO-HAND COORDINATION

X1 Tracking Error

Cumulative tracking error for the X1 axis (horizontal displacement of the cross from the target) for the final 2 minutes of the test period.

Y1 Tracking Error

Cumulative tracking error for the Y1 axis (vertical displacement of the cross from the target) for the final 2 minutes of the test period.

COMPLEX COORDINATION

X2 Tracking Error

Cumulative tracking error for the X2 axis (horizontal displacement of the cross from the center of the screen) for the final 2 minutes of the test period.

Y2 Tracking Error

Cumulative tracking error for the Y2 axis (horizontal displacement of the cross from the center of the screen) for the final 2 minutes of the test period.

Z2 Tracking Error

Cumulative tracking error for the Z2 axis (displacement of the rudder bar from the center point at the bottom of the screen) for the final 2 minutes of the test period.

ENCODING SPEED

Several tests in the BAT battery rely on response latencies as an indicator of test performance. The standard scoring technique for tests of this type uses data only from trials that were answered correctly when computing summary scores. For the BAT battery, this procedure includes the following tests and scores used to evaluate performance:

Encoding Speed Test

- Average Response Time: Average response time in milliseconds based on all trials answered correctly.

- Percent Correct: Average Response Time x Percent Correct. This is a response time by percent correct interaction term: (Subject's average response time-grand mean average response time) x (Subject's percent correct -grand mean percent correct).

Mental Rotation

- Average Response Time: Average response time in milliseconds based on all trials answered correctly.

- Standard Deviation: Standard deviation of response time in milliseconds based on all trials answered correctly.

- Percent Correct:

Item Recognition

- Average Response Time: Average response time in milliseconds based on all trials answered correctly.

- Slope: This score represents a regression slope for the best-fitting line for average response time to digit strings of differing lengths (1, 2, 3, 4, 5 or 6 digits).

- Intercept: This score represents a regression intercept for the best-fitting line for average response time to digit strings of different lengths (1, 2, 3, 4, 5 or 6 digits).

- Percent Correct:

Time-Sharing

- Slope (Tracking Difficulty): Average tracking difficulty was computed for each subject during each minute of the practice trials (minutes 1-10). This score represents a regression slope based on the best-fitting line describing the eight average tracking difficulty scores for minutes 3-10 of this test (learning rate on the tracking task).

- Intercept (Tracking Difficulty): This score represents a regression intercept based on the best-fitting line describing the eight 1-minute average tracking difficulty scores for minutes 3-10 of the test.

- Average Tracking Difficulty: Average tracking difficulty achieved during minutes 11-19.

- Average Response Time: Average response time in milliseconds to cancel the digits that appear during the dual-task trials (minutes 11-16).

- Average Response Time x Tracking Difficulty: This is a response time by tracking difficulty interaction term based on performance during the dual-task trails (minutes 11-15).

Self-Crediting Word Knowledge

- Average Response Time: Average response time in milliseconds based on all trials answered correctly.

- Percent Correct:

- Average Response Time x Percent Correct: This is a response time by percent correct interaction term.

- Bet: This score is the average bet the subject made before block of trials. It reflects self-confidence regarding expected performance. Higher scores reflect greater self-confidence.

Activities Interest Inventory

- Number of High-Risk Choices: This score is the number of high-risk choices made by the subject.

- Average Response Time: Average response time across all 81 trials.

SUMMARY

The BAT, was developed by Illiana Aviation Sciences, Inc. and Technical Solutions, Inc. for the USAF Human Resources Laboratory. The battery originally consisted of 15 tests, but successive studies have yielded a subset of nine tests that provide the greatest gains in predictive validity when combined with the (AFOQT).

The nine subtests are:

1. Test Battery Introduction (Biographical Information, e. g.,age, handedness, previous flying experience).

2. Two-Hand Coordination (rotary pursuit).

3. Complex Coordination (stick and rudder).

4. Encoding Speed (Verbal classification).

5. Mental Rotation (Spatial transformation and classification).

6. Item Recognition (Short-term memory, storage, search, and comparison).

7. Time Sharing (Higher order tracking ability, learning rate and time sharing ability).

8. Self-Crediting Word Knowledge (self-assessment ability, self-confidence).

9. Activities Interest Inventory (survival attitudes).

It should be noted which tests were removed from the BAT due to their low predictive validity since some similar tests remain in other test batteries. Two of these tests, Decision-Making Speed, and the Automated Air Crew Personality Profiles have been modified and currently are being evaluated for inclusion in the BAT battery.

The deleted tests are:

Dot Estimation

Two boxes containing an arbitrary number of dots are presented simultaneously on the screen. Cne box contains one more dot than the other. The subject's task is to decide, as quickly as possible, which box has the greater number of dots. Subject are not told to count the dots in each box. They are only instructed to decide as quickly and accurately as possible which box has the greater number.

Reaction time and accuracy of response are recorded on each trial. This is the only test in the battery that has a fixed time limit (5 minutes, for a maximum of 55 trials). The psychological factor assessed by this test is impulsiveness/decisiveness.

Digit memory

A string of four digits is presented simultaneously and in random order. The subject is instructed to enter the digit string on a data entry keypad in the same order as presented. In addition to the recording of response accuracy and overall response time, a measure of perceptual speed is taken by forcing the subject to press a special ENABLE key that activates the data entry keypad buttons on each trial.

The most important attribute measured by this test is perceptual speed. There are 20 trials, requiring approximately 5 minutes to complete.

Immediate/Delayed Memory

In this test, a sequence of digits is presented, and the subject is required to respond by indicating the digit that occurred either one or two digits previously. The one-back and two-back subtests have two parts. In the first part, the digits are presented for 1/2 second, followed by a 2-second inter-stimulus interval. In the second part, the inter-stimulus interval is 5 seconds. Thus, for both subtests, part one deals with "immediate" memory and part two with "delayed" memory.

There are 25 trials in each subtest (one-versus two-back) for each length of latency condition (2 versus 5 seconds) resulting in 100 trials. As with the other tests, response time and accuracy are recorded on each trial.

This test assesses continuous short-term memory storage and retrieval operations.

Decision-Making Speed

In this choice-reaction-time test, one of several alternative signals is presented to the subject. The subject is required to respond to the signal as quickly as possible. The critical manipulation in this test is the amount of uncertainty that must be resolved to make the response decision. As increasing numbers of potential alternatives are increased, greater uncertainty exists and the decision is made more slowly. This test consists of four subtests.

In subtest one, the subject knows both where and when a signal is to occur; in subtest two, the subject knows where but not when; in subtest three, when but not where; and finally, in subtest four, the subject knows neither where nor when. Each subtest has three parts. In part one, two potential signals and responses are defined. There are four potential signals and responses in part two and eight potential signals and responses in part three. Therefore, degree of uncertainty of signal is manipulated in three ways: location of occurrence, time of occurrence, and number of signals/responses. There are 12 trials within each part of each subtest, resulting in 144 trials (12x3x4). Response time and accuracy of response (correct/incorrect) are recorded for each trial.

The Decision-Making Speed Test assesses a variety of psychological factors. These include simple choice reaction time under varying degrees of information load and spatial and temporal uncertainty, as well as low-level cognitive and high-level sensory-perceptual motor involvement.

Risk-Taking

Ten boxes are presented in two rows of five boxes each. The subject is told that 9 of the 10 boxes contain a reward, and one box is a "disaster" box. If the selected box contains a payoff, the subject is allowed to keep it; but if the subject chooses the disaster box, all of the payoff earned on that trial is lost. The average number of boxes selected provides an index of the subject's tendency to take risks when making decisions.

Response time per choice and number of boxes chosen are recorded on each of 30 trials. Unknown to the subject, there is no "disaster box" (i.e., no risk) for 12 of the 30 trials. This method was used to achieve a uncontaminated measure of risk-taking behavior, since performance on the "disaster box" trials could be affected by chance.

Embedded Figures

A simple geometric figure and two complex geometric figures are presented to the subject. The subject's task is to decide which of two complex figures has the simple figure within it and to press the button corresponding to the figure. Speed and accuracy of response

are recorded on each of 30 trials. This test is designed to assess the psychological factor of field dependence/independence.

Automated Air Crew Personality Profiles

This is a questionnaire that examines the subject's attitudes and interests. The subject is given 66 questions, each requiring a choice between two alternatives. The subject is instructed not to spend time pondering responses, but to give the first answer that comes to mind. The questionnaire is a traditional personality inventory specially targeted for air crew selection.

CONCLUSIONS

Carretta (1990) concluded that a combination of the AFOQT and the BAT demonstrated sufficient utility for supporting USAF pilot candidate selections. The USAF plans to place the current version of the BAT in operation in the spring of 1992.

REFERENCES

Carretta, T. R. (1987a). Basic Attributes Tests (BAT) system: Development of an automated test battery for pilot selection. (AFHRL-TR-87-9) Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T. R. (1987b). Time sharing ability as a predictor of flight training performance (AFHRL-TP-86-69, AD-A181 838). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T. R. (1988). Relationship of encoding speed and memory tests to flight training performance (AFHRL-TP-87-49, AD-A190 670). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T. R. (1989). USAF pilot selection and classification systems. Aviation, Space, and Environmental Medicine, 60, 46- 49.

Carretta, T. R. (1990a). Basic attributes test (BAT): A preliminary comparison between reserve officer training corp (ROTC) and officer training school (OTS) pilot candidates. (AFHRL-TR-89-50) Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T. R. (1990b). Cross-validation of experimental USAF pilot training performance models. (AFHRL-TR-89-68) Brooks AFB, TX: Human Resources Directorate Manpower and Personnel Division.

Carretta, T. R. (1991). Short term test-retest reliability of an experimental version of the Basic Attributes Test Battery. (AL-TP-1991-0001) Brooks AFB, TX: Human Resources Directorate Manpower and Personnel Division.

Carretta, T. R., & Siem, F. M. (1988). Personality, attitudes and pilot training performance: Final analysis (AFHRL-TP-88-23, AD-A199 983). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Imhoff, D. L., & Levine, J. M. (1981). Perceptual-motor and cognitive performance task battery for pilot selection (AFHRL-Kantor, J. E. (1984). The USAF pilot selection and classification research program. Aviation, Space, and Environmental Medicine, 56, 258-261.

Mercatante, T. A. (1988). The reliability and validity of psychomotor aptitude for pilot selection. Unpublished master's thesis, St. Mary's University, San Antonio, TX.

TABLE 1
Reliability Estimates for BAT Performance Scores[**]

Test score	N	No of Scores[*]	Cronbach's alpha	Reference
Two-Hand Coordination: X1 Tracking Error	233	10	.94	Mercatante, 1988
Complex Coordination:				
X2 Tracking Error	233	10	.95	Mercatante, 1988
Y2 Tracking Error	233	10	.99	
Z2 Tracking Error	233	10	.94	
Encoding Speed:				
Response Time	2,219	96	.96	Carretta, 1988
Response Accuracy	2,219	96	.71	
Mental Rotation:				
Response Time	1,685	72	.97	Carretta, 1990
Response Accuracy	1,685	72	.90	
Item Recognition:				
Response Time	1,500	48	.95	Carretta, 1990
Response Accuracy	1,500	48	.54	
Time-Sharing:				
Tracking Difficulty	1,130	19	.96	Carretta,1987a
Self-Crediting Word Knowledge:				
Response Time	1,992	30	.89	Carretta & Siem,
Response Accuracy	1,992	30	.65	1988
Activities Interest Inventory:				
Response Time	1,992	81	.95	Carretta & Siem,
Response Choice	1,992	81	.86	1988

[*] For the Two-Hand Coordination Test, Complex Coordination Test and Time-Sharing Test, "Scores" refer to summed tracking performance over time. For the other tests, "Scores" refer to test items.

[**] Carretta, 1990

COMPLEX COGNITIVE ASSESSMENT BATTERY
(CCAB)

PURPOSE

The Complex Cognitive Assessment Battery (CCAB) was developed to measure higher order cognitive processes involved in high level military decision making. The battery was to be used to evaluate the cognitive performance effects of prophylactic drugs used in chemical defense. Since many cognitive skills required to perform the tests are characteristic of those required for piloting aircraft, the CCAB has been investigated to assess flying potential.

DESCRIPTION

The CCAB was developed by the Analytical Assessments Center (AAC) of the Eaton Corporation for the Army Research Institute. The CCAB is composed of eight sub-tests, each designed to measure more than one cognitive process. Each sub-test includes a set of instruction screens, practice problems, a quiz to evaluate the subjects' ability to understand the tasks, and a set of test problems. The sub-tests are as follows.

TOWER PUZZLE

Purpose

The Tower Puzzle test is a computer-administered version of the "Tower of Hanoi" problem. It was developed to measure quantitative reasoning, planning, situation assessment, decision making, and problem solving functions. In addition, the Tower Puzzle test assesses perception of form, concept formation, communication, and creativity.

Description

A monitor displays three posts with blocks stacked on each post. Five blocks are distributed among the three posts. The goal is to create a single "tower" of blocks on the center post, with the largest blocks at the bottom of the posts and the smallest at the top.

All practice and trial problems have five blocks stacked on three separate posts. The problems differ in the initial distribution of the blocks among the three posts. This permits test-retest of subject performance. The following are descriptions of the sub-tests.

FOLLOWING DIRECTIONS

Purpose

The Following Directions test was designed to measure high-level cognitive abilities concerned with responding to data and going beyond the data. It tests attention to detail, memory retrieval, time sharing, comprehension, verbal and quantitative reasoning, and communication.

Description

Directions are displayed on the monitor. The cursor is moved by pressing the RIGHT, LEFT, UP and DOWN keys on the numeric keypad on the computer keyboard. The goal is to follow the directions as accurately and rapidly as possible.

The following directions task includes three general procedures. The first to underline, or "mark," words as instructed. The second is to remove some underlining, or "unmark," works as instructed. The third procedure is to mark and unmark words based on values of X, Y, and Z, which are given at the bottom of the display. These values of X, Y, Z change as each new trial is presented.

WORD ANAGRAMS

Purpose

The Word Anagram test is a computer-administered variation of the well known anagram problem, that extensively investigated in the experimental psychology literature. This adaptation of the anagram task involves constructing a word or multiple words from a series of scrambled letters. This test is intended to measure attention to detail, perception of form, memory retrieval, concept formation, planning, decision making, problem solving, and creativity.

Description

In the classic anagram task, the subject is given a set of scrambled letters and is required to use all the available letters to construct one solution word in an allotted time. For example, the subject may be presented with the letters U T R I F where the solution would be FRUIT. In the present variation six unique letters are displayed in the upper portion of a monitor. The letters are made up of four consonants and two vowels, for example, H A L T M I. The anagram problem in this task is to construct as many 3 and 4 letter words as possible from the displayed letters within an allotted time. Each time period makes up an individual test trial.

In this task, the subject selects letters from those displayed (e.g., H A L T M I) to form words of 3 letters (e.g., aim, hit, etc.) or 4 letters (e.g., malt, hilt) by first moving the cursor to the desired letter and then pressing the space-bar to select the letters one at a time. Any of the letters may be used more than once in forming a given word: for example, the words "all" and "hill" would be acceptable words created from the given displayed letters. As the subject creates words, the letters remain displayed at the top of the monitor and the words that are formed are displayed in a numbered list in the middle of the monitor.

LOGICAL RELATIONS

Purpose

The Logical Relations test was designed to measure the ability to comprehend a pair of premises that defines a logical relationship among three items. The subject must decide which item has the greatest or least value (logical syllogism). Performance reflects the combined

cognitive skills of attention to detail, memory retrieval, comprehension, verbal and quantitative reasoning, and problem solving.

Description

The first task on each trial is to examine two premises that depict relations (e.g., worse than/better than, longer than/shorter than, faster than/slower than, or 'ighter than/darker than) among three items (A, B, and C). For example the two premises might be "B is lighter than A", and "B is darker than C." Given these two premises each trial is proceeded with a question displayed on the monitor regarding the item with either the largest ("darker") or smallest ("lighter") value. For example, the question that might follow from the two premises above is "Which is darkest?" The subject responds by entering the letter A, B, or C. The total time allowed for a given trial is set by the Test Administrator. Since a trial is made up of several individual syllogisms, the subjects can attempt to solve the syllogisms at their own pace. The objective is to solve all syllogisms for a given trial within the time limit.

MARK NUMBERS

Purpose

The Mark Numbers test was designed to measure the ability to search an array of numbers and make specified quantitative decisions about each number. The subject also must periodically respond to a secondary task involving a flashing command statement, that directs the subject to select one of two flashing numbers in the array. Performance on this test reflects principally the combined cognitive skills of attention to detail, memory retrieval, time sharing, comprehension, verbal and quantitative reasoning, and situation assessment.

Description

The task on each trial is to examine a 6 x 10 array of two-digit number and to underline any number that falls within the category specified in an instruction presented at the bottom of the display (e.g., "Mark ODD numbers between 15 and 45"). Subjects mark the numbers by using the arrow keys to move to each number that meets the given criteria and then the press the space, that marks the number by placing an underline beneath the number.

Besides the primary task of marking the numbers at the bottom of the screen, a secondary task begins twenty seconds into the trial. This secondary task is made up of two numbers in the array that begin flashing on and off. These two numbers are always opposite of the number type specified in the primary task. If the primary task involves marking ODD numbers, then the flashing numbers will be EVEN numbers, and vise versa. When they begin flashing, a second instruction flashes simultaneously with the two flashing numbers at the bottom of the display. This instruction specifies which of the two flashing numbers is to be marked (e. g., "Mark SMALLER flashing number" or "Mark BIGGER flashing number"). The subject has ten seconds to mark the specified flashing number.

When the correct flashing number is marked, or when ten seconds have elapsed, the flashing stops and the subject proceeds with the primary task of marking numbers in the array

as specified by the non-flashing, original instruction. Later, another pair of numbers flashes with a new instruction appearing approximately every twenty seconds. This secondary task will interrupt the primary task throughout the remainder of the trail.

The total time taken for the test can be left open, allowing subjects to progress at their own pace, or the total time for the test can be set at 3, 6 or 9 minutes. The time allotted per trial ranges between 60, 90, and 120 seconds, and can be established by the test administrator. When all the two-digit numbers specified by the primary instruction have been marked, the trial is completed. The subject is instructed that points will be awarded only for accurate marks and points will be deducted for inaccurate marks.

NUMBERS AND WORDS

Purpose

The Numbers and Words test was designed to measure the ability to perform two cognitive tasks simultaneously. The test employs a classic time sharing paradigm, involving a perceptual monitoring task and a running memory task. In addition, to measuring time sharing performance, the task is included to assess the cognitive functions of attention to detail, perception of form, memory retrieval, time sharing, concept formation and decision making.

Description

In the Numbers and Words test two tasks are presented simultaneously compelling the subject to share time between them. The "numbers" task involves a running memory task where a previously displayed number must be recalled as each number is displayed. The "words" task requires the subject to identify, when possible, a 3-letter word that emerges gradually over the course of a trial from a white blob to an easily discernible word. As the word emerges, the subject must monitor it; and simultaneously, respond to the continuous stream of displayed numbers. The display is divided into 3 panels; two at the top and one across the bottom. The upper right panel is used to present a 3-letter word, that emerges from a white rectangle slowly overtime. The full emergence of the word is accomplished as more "noise" in the form of dots (pixels) are removed, until the letters forming the word are displayed in virtual clarity.

The bottom panel presents 12 alternative words from which the subject selects a word that appears to match the emerging word displayed in the upper right panel. These alternative words are labeled 'A' through 'L'. On each trial, the subject is allowed two tries to identify the correct word. Feedback is given on the monitor if the first guess is correct. If it is incorrect, the word in the upper right panel continues to emerge until the second guess is attempted.

The upper left panel presents a series of numbers one at a time, throughout the timed trial segments. The number displayed changes every 4 seconds or sooner. Each new number is displayed for up to 2.5 seconds or until the subject responds with the previous number. After the new number is removed, there is a fixed delay of 1.5 seconds. When the number changes, the subject's task is to press the number key 1, 2, or 3 to indicate the previous number in the series. Thus, this second task is one of "running recognition." and serves to create a time-sharing situation.

INFORMATION PURCHASE

Purpose

The Information Purchase test was designed to measure the ability to decide when sufficient information has been gathered to identify a target, in converging multiple trials. Performance reflects not only the combined cognitive skills of attention to detail, perception of form, memory retrieval, time sharing, concept formation, quantitative reasoning, situation assessment, and decision making, but also planning and problem solving.

Description

The basic stimulus display for this test is a grid that is 14 columns wide and 6 rows high. The rows are labeled A through F, with the letters placed both on the right and left side of the grid. The test begins with a single digit displayed in the first column of each rows. Only the digits, 0, 1, and 2 are used.

The subject can fill additional rows with numbers by pressing the space-bar. Each time the bar is pressed, the next column of all the rows will be filled with a number (0, 1, or 2). The numbers are selected pseudo-randomly. The task is to decide in which row the sum of the numbers will be the largest if all columns were filled. Though 14 columns are displayed and may be filled with numbers, each row is designed as a 20 column row. Because of this, even if all 14 columns were filled, the subject would still be "predicting" which row would end with the highest total if all 20 columns were filled with numbers.

High scores are achieved by identifying the row quickly, with as few columns filled in with numbers as possible. When subjects make a decision, they enter the corresponding letter (A through F). Subject are allowed only two guesses. The more information purchased (i.e., the more rows the subject filled with numbers by pressing the space-bar), the lower the point score when a correct guess is made. The total time taken for the test can be set at 3, 6, or 9 minutes.

Subjects are allowed two guesses to predict which row will end up with the highest total. More points are awarded for a correct guess that is made earlier in the trial with less information. If the subject predicts incorrectly, a beep tone will sound and an error message will be displayed. When the correct prediction is entered, time has run out, or both guesses have been expended, the correct row letter is displayed.

ROUTE PLANNING

Purpose

The Route Planning test was designed to measure the ability to plan and execute a route from a starting position to an ending (target) position. Simple rules of movement with a variety of constraints must be followed. Performance should reflect the combined cognitive skills of planning and communication, and others such as perception of form, attention to detail, situation assessment, decision making, concept formation, quantitative reasoning, and problem solving.

Description

A 5 x 5 matrix (25 squares), with 11 squares shaded-in is presented, determined by pseudo-random selection. The remaining, unshaded squares display 14 letters of the alphabet. Since the matrix contains 25 squares, all letters of the alphabet, except the letter Z, could be displayed. However, the 11 shaded squares block out 11 of the letters so that only 14 letters are visible at any given time. The letters are ordered from left to right in the matrix, with the top left square containing the letter A, and the bottom right square of the last row containing the letter Y.

The task is to move from a designated starting square to a designated ending square by planning and communicating a route. Shaded squares can be traversed, but cannot be landed on. Moves are restricted to two "L" shaped forms, either "one out and two over" or two out and one over." This move pattern in similar to the "Knight's" move in a game of chess, though the subject does not need to be familiar with chess to perform the test. The subject indicates the moves by entering the letters of the squares to be moved to in sequence.

A trial ends when the ending square is reached or when the allotted time (either 60, 90, or 120 seconds as predetermined by the experimenter) has elapsed. A complete test session consists of 2 to 4 practice problems, and 3 to 9 individual test trials; these parameters are also under the experimenter's control. The subject is instructed that more points will be awarded for solutions that are obtained quickly and with the minimum number of moves.

PUBLISHER/AGENCY

U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

ADMINISTRATION

Time

An "open" setting is used in most testing situations. The test is completed when a specified number of trials is finished.

Apparatus

An IBM compatible XT or AT class microcomputer with a minimum of 384K memory, a hard drive and color graphics monitor.

Procedure

The test is self administered once a test session is set up and the subject begins. Each test has a set of instruction screens, practice problems, a quiz to measure the subject's understanding of the tasks, and a set of test problems that can be used for repeated testing. The

software allows the administrator to vary the test sessions to select which tests to be given and in what order, and if the test instructions and a quiz are to be given.

Reliabilit,

See Table 2.

Criteria

Not available

Validity

Not available

Norms

Not available

Scoring

The test is automatically computer-scored and a set of scores are stored on disk for later printout.

SUMMARY

The following is brief summary of the eight sub-tests. Each sub-test measures multiple complex cognitive capabilities.

1. Tower Puzzle - planning, situation assessment, problem solving, decision-making, perception of form, concept formation, quantitative reasoning, communication, and creativity.

2. Following Directions - attention to detail, memory retrieval, time sharing, comprehension, verbal reasoning, quantitative reasoning, and communication.

3. Word Anagrams -memory retrieval, creativity, perception of form, planning, decision-making, attention to detail, concept formation, and problem solving.

4. Logical Relations-comprehension, verbal reasoning, quantitative reasoning, attention to detail, memory retrieval, and concept formation.

5. Mark Numbers - attention to detail, quantitative reasoning, time sharing, memory retrieval, comprehension, verbal reasoning, and situation assessment.

6. Numbers and Words - perception of form, time sharing, attention to detail, memory retrieval, concept formation, and decision making.

7. Information Purchase -situation assessment, decision making, attention to detail, quantitative reasoning, planning, perception of form, memory retrieval, time sharing, concept formation, and problem solving.

8. Route Planning -planning, perception of form, situation assessment, communication, problem solving, attention to detail, concept formation, quantitative reasoning, and decision making.

● Advantages

-- The CCAB has the potential to measure higher order cognitive processes involved in complex decision making that may not be measured by other batteries.

-- A comprehensive users guide is available from the US Army Research Institute.

● Disadvantages

-- Researchers are experiencing software problems with the current version.

-- Validity studies and normative data are not available to date.

REFERENCES

Geiselman, R. E., & Samet, M. G. (1986). Complex Cognitive Assessment Battery (CCAB): Pilot test results. (AAC-TR-33212) Alexandria, VA: U.S. Army Research Institute.

Samet, M., Marshall-Miles, J. C., & Albarian, G. (1987). Complex Cognitive Assessment Battery (CCAB): Final test administrator user guide. (AAC-UM-33212) Alexandria, VA: U.S. Army Research Institute.

Samet, M., Geiselman, R. E., Zajaczkowski, F. & Marshall-Miles, J. (1986). Complex Cognitive Assessment Battery (CCAB): Test descriptions. Alexandria, VA: U.S. Army Research Institute.

TABLE 2

Test-retest Reliability Coefficients***

(Number of Pair of Measures Available are Shown in Parentheses)

Word Anagrams	.89**	(53)
Tower Puzzle	.74**	(39)
Logical Relations	.68**	(54)
Following Directions	.95**	(30)
Mark Numbers	.66*	(18)
Route Planning	.85**	(51)
Weighted Average Across Tests	.80	

** p < .001

* p < .01

*** Geiselman and Samet, 1986

MICRO-COMPUTERIZED PERSONNEL APTITUDE TESTER
(MICROPAT)

PURPOSE

The MICROPAT was originally designed for assessing suitability for Army rotary-wing pilot training. It was planned to be a self contained unit that would be easily administered and scored.

DESCRIPTION

Initial research and development on MICROPAT began in 1980 and was conducted for the British Army Air Corps whose major operational interest was helicopters. In 1985, The Royal Navy took over funding for the project. MICROPAT is an automated testing and record-keeping system for air crew selection. The current version (5.0) contains two categories of tests --psychomotor ability and information management ability.

The following is a description of the nine sub-tests:

COMP2D (10 minutes)

This is a two-dimensional compensatory tracking task using a position control law where the X-axis is controlled by foot-pedals and the Y-axis by a joystick. Three 2 minute trials are presented and error scores for each axis are taken. The six basic errors scores are highly correlated and so a composite error score is taken.

ADTRACK2 (10 minutes)

This is a one-dimensional pursuit tracking task with joystick control. There are three trials with 27 five second intervals (3 times 2 minutes 15 seconds). There are 17 levels of difficulty and the computer measures the candidates' performance every 5 seconds to vary or maintain the level of difficulty.

The basic score is the mean difficulty level over the last 11 periods (5 seconds each) of each trial. Two measures are then derived: the mean level across the 3 trials and the difference between the first and last trial levels.

ADTRACK3 (10 minutes)

This is similar to ADTRACK2 except that the candidate can control the adaptation. Difficulty level is increased or decreased by pressing a key. Two trials of 2 minutes are administered. The level of difficulty reached is measured for each second of the last minute of each trial. Three measures are derived; the average level in the last minute of each trial, the mean number of key presses for the 2 trials and the average tracking error across both trails.

LANDING (15 minutes)

This is a 3-dimensional tracking task that resembles an aircraft landing simulator. The candidate controls an aircraft symbol with a joystick and sliding power control. Information about height, speed and fuel reserve is provided. This test examines "division of attention" between the "within-cockpit" information (height, speed, and fuel) and the "external" visual information about attitude and position. Subjects are given one practice flight followed by 5 assessed flights. Retest scores show there is a practice effect.

Three basic measures are taken: accuracy of final landing alignment, speed control (mean speed minus the final speed) and rate of descent (height descended divided by fuel used). From these basic measures, 6 final measures are computed: mean accuracy over the last 3 flights; difference in accuracy between first 2 and last three flights; mean speed control over last 3 flights; difference between mean speed control on first 2 and 3 flights; mean rate of descent on last 3 flights; difference between rate of descent on first 2 and last 3 flights.

DUALTASK (20 minutes and SUBTRACT (15 minutes)

This test was designed to measure division of attention. The candidate is first presented with 30 subtraction problems that must be checked as correct or incorrect. The next section of the test involves a two-dimensional step-tracking task with laterally-inverted position control; the subject must make a box jump to 31 different locations on the screen. In the final section, the subject must check 25 subtraction problems and move the box 31 times.

The current measures are: total time required to check the last 25 problems divided by the number correct; percent of accuracy in last 25 problems; mean per "jump" in last 25 jumps in the tracking only condition; percent of change in the time taken to complete problems in the 2 conditions with a correction for wrong answers; change in accuracy in the problems in the 2 conditions; percent of change in tracking time in the 2 conditions.

PLANE

This test involves an application of signal detection theory. Two aircraft silhouettes are displayed, that vary in size. Candidates must decide if an enemy (large silhouette) or friendly aircraft (small silhouette) is on the display within 1.5 seconds. There are 3 trials of 50 silhouettes with discrimination difficulty progressively increasing. Trials of earlier versions of SIGNAL showed that decision speed and accuracy measures had good psychometric properties but signal detection type parameters did not. The basic measures used in developing PLANE are: mean decision response time; probability of a hit averaged over the 3 test trials; probability of a false alarm averaged over the 3 trails; measure of discriminative sensitivity averaged over the 3 trials; log likelihood ratio-a-measure of response bias or the tendency to respond/not respond averaged over the 3 trials.

SCHEDULE

Five columns are displayed, each containing a target box of varying values (1,2,3 etc.). When a key is pressed for a column, a line begins to grow toward the target; only one column

can be active at a time: if another key is pressed, the first column is canceled and a new line starts in the new column. Boxes have a limited life and lines grow at different rates in different columns. Also, a random variable is added to the box life and the line growth rate. The candidate is instructed to achieve the highest score as possible (i.e., hits on boxes). After 2 short practice trials, there are 4 trials of 128 seconds. The computer applies an optimal strategy algorithm and calculates the candidate's score. The candidate's performance is expressed as a percentage of an optimal score. The 2 measures taken are: the mean percent score across all 4 trials and the increase in mean percent score from first 2 to last 2 trials.

RISK

This is a gambling task where the candidate may select up to 8 keys; one key is a penalty key and results in the loss of any points gained from previous selections. In the first set of trials there is a penalty key on each trial but in the second set the penalty key operates only 50% of the time. The test taps relative cautiousness-risk taking and the adaptation of strategy from first to second set of trials.

The current version uses the scenario of gaining valuable documents from 8 locations, with an enemy ambush at one of them. Four blocks of trials are presented, 2 blocks in each condition mentioned above (Blocks 1 and 4 penalty key 100% of time, Blocks 2 and 3, 50% of time). The measures are: mean number of key presses in the 100% condition; proximity to optimal number of key presses in 100% condition; mean number of key presses in 50% condition; difference between 100% and 50% conditions; difference in number of keys pressed following a win versus following a loss in 100% condition but for 50% condition.

NAVOR

This was designed as a measure of spatial orientation and was designed for helicopter Observers. The task is essentially dead-reckoning: candidates have to fly a journey mentally and then find their home base. The flight instructions tell them to fly a given distance on a given heading. The complexity of the task can be varied by the number of legs flown (2 or 4) and the number of compass points (4 or 8). The measures recorded are: the mean final heading error; the mean distance error; inspection time needed for the flight instructions and; the time taken to indicate position of home base after acknowledging last instructions.

MANIKIN

This a common cognitive paradigm. This version was originally developed as a marker variable for the NAVOR and NAVCALC trials. The screen displays a male figure holding an object in one hand; the figure may be (1) upright, (2) upside down or (3) lying down, facing the candidate or with his back toward him. The subject must decide if the object is in the right or left hand. There are 3 x 24 trials with feedback. A principal components analysis of all the measures used in the trial produced 3 components: Reaction Time (44% of variance), Percent Correction (22%) and a third unnamed measure (15%).

<u>NAVCALC</u>

This is a work sample type test designed to simulate one aspect of Observer training/producing and updating a flight log.

The candidate must copy a log from the screen. The log contains several locations on a flight plan, their radio frequency, departure times, turnaround times, and flight times between locations. The computer updates some items on the log as the test progresses and the candidate must change the log accordingly. Questions about the log are interspersed with the log update information. If the subject gives an incorrect answer, the correct answer is displayed and he can amend his log accordingly.

PUBLISHER/AGENCY

Ministry of Defence (Navy)
London, U.K.

ADMINISTRATION

<u>Time</u>

Testing time required is approximately 110 minutes.

<u>Apparatus</u>

The MICROPAT battery requires no specially designed equipment and will run on an AT class microcomputer with a color monitor. This feature would allow the use of a portable system.

RELIABILITY

Internal consistency estimates of sub-test scores ranged from .53 to .93 (Jones and Abram, 1988).

APPARATUS

The test requires no specially designed equipment. MICROPAT will run on an AT class microcomputer with a color monitor. This feature would allow the use of a portable system.

CRITERIA

The Royal Navy (RN) trial of MICROPAT used four major criteria: (1) Completion of Pilot Elementary and Basic Flying Training (obtained Pilot "wings"); (2) Successfully passing through elementary/Basic/Advanced/Operational Pilot training; (3) Passing Observer Basic Training (BOC); and (4) successfully passing through Basic/Advanced/Operational Observer training.

VALIDITY

Prediction of RN Pilot Elementary-Basic Rotary flying Training: A step wise multiple regression analysis of all MICROPAT variables yielded a correlation of $r = .51$ with training outcome.

Prediction of RN Pilot Elementary-Advanced Rotary Flying Training: A multiple correlation of $r = .29$ was found when corrected for restriction of range.

Prediction of RN observer Basic-Operational Flying Training: A multiple correlation of $r = .55$ (uncorrected) with successful pass/fail through Basic, Advanced and Operational phases of Observer training was reported. These correlations were based on a sample of 109 pilots and 111 observers (Bartram, 1990).

NORMS

Current norms are based on sample sizes between 11 and 859. "T" scores and percentiles are available. The subjects were observer and pilot rotary wing flying training candidates. Norms are currently being gathered on air crew applicants for Cathay Pacific Airlines.

SCORING

Version 5 (MS-DOS version) runs under a commercially available administration marketed by ASE, NFER-NELSON ("SCREENTEST"). An interpretation package is under development to calculate composite predictor scores and generate a narrative to provide feedback of results.

SUMMARY

The nine subtests are:

1. Landing (superficially resembling an aircraft simulator).

2-4. Adtrack-1, 2 and 3 (three tasks of one dimensional pursuit tracking using a joystick controller.

5. Risk (a gambling task involving explicit decision making where the costs and payoffs are made known).

6. Signal (a signal detection task involving implicit decision making. Sensory information must be interpreted and some correction to the flight path must be made).

7. Schedule (sorting priorities).

8. Dual Task (two-dimensional step-tracking and mental arithmetic).

9.   Memory Span (an adaptive test of digit-span), Bartram, 1986.

● Possible Advantages

-- High face validity
-- Minimum hardware requirements
-- Portability
-- Output files can be read by current DOS based data analysis packages, e. g., SPSS-PC+, and Lotus 123.
-- The battery has been refined and validated as a composite test used as a composite test.
-- Reasonable time required for administration.

● Possible Disadvantages

-- Developed for selection of rotary wing combat pilots and to date validity studies have not been reported for fixed-wing pilots. Studies are in progress.

REFERENCES

Bartram, D. (1986). Development and evaluation of MICROPAT Version 4, (TR-184) London: Ministry of Defense, and University of Hull, U.K.: Ergonomics Research Group.

Bartram, D., & Marshall, L. (1990). Validation of MICROPAT Version 4, (TR-248), London: Ministry of Defense, and University of Hull, U.K.: Ergonomics Research Group.

Jones, A., & Abram, M. (1988). MICROPAT databook: Issue two, (TR-255), London: Ministry of Defense.

NAVAL AVIATION SELECTION TEST BATTERY

PURPOSE

The Naval Aviation Selection Test Battery was intended to augment the Navy's paper-and-pencil selection test for aviators.

DESCRIPTION

The Naval Aviation Selection Test Battery is an automated performance-based test battery. Several combinations of predictors have been studied over the years. The latest study (Delaney, 1990), includes the automated dichotic listening task (DLT) and psychomotor tasks (PMT). On the DLT, subjects listen to two different series of letters and numbers presented simultaneously to their ears over binaural headphones. The PMT includes single or dual modes of stick/rudder/throttle tasks.

PUBLISHER/AGENCY

Naval Aerospace Medical Research Laboratory.
Naval Air Station
Pensacola, Florida 32508-5700

ADMINISTRATION

Time required: Approximately 75 minutes including brief breaks.

APPARATUS AND PROCEDURE

All tasks are presented on an Apple IIe computer.

PSYCHOMOTOR TASKS

On the PMT subjects are required to maintain first one, then two, and finally three randomly displaced cursers on fixed targets on a CRT by manipulating joysticks and foot pedals. Psychomotor test scores are the cumulated total absolute errors from the target in pixels.

DICHOTIC LISTENING TEST (DLT)

On the DLT, subjects listen to two different series of letters and numbers presented simultaneously to each ear over binaural headphones. Subjects are instructed which ear to attend to on each trial. Subjects are told to indicate the digits presented to the designated ear in the order of their occurrence. Responses are entered with the left hand on a keypad placed in front of the subjects.

DUAL TASKS

For the dual tasks, subjects are asked to simultaneously perform a 4.5 - minute PMT and a 12-trial DLT. Performance is scored in the same way as for the single tasks.

RELIABILITY

Not available

CRITERIA

Grades in the academic portion of naval flight training and pass/fail in the flying portion of primary flight training.

VALIDITY

The DLT/PMT tasks were administered to 677 student aviators upon completion of the academic portion of naval flight training and waiting to begin primary flying training. Correlational analyses showed primary flight grades were significantly related to PMT scores ($r = -.26$ to $-.41$, $P < .05$), and DLT scores ($r = -.22$ to $-.28$). Multiple regression analysis yielded a validity coefficient for a combination of performance measures of $r = .442$. The 19.5% of the flight grade variance accounted for by the performance-based tests was largely independent of the 16.6% variance accounted for by a combination of current selection tests and demographic variables. Individual performance variables were not significantly related to the pass/fail criterion (Delaney, 1990).

NORMS

Not available

SCORING

Not available

SUMMARY

- ● Advantages

  -- No specially designed equipment is required.

  -- All tasks are presented on an Apple IIe computer.

- ● Disadvantages

  -- None of the tasks were significantly related to the pass/fail performance criteria.

A-35

REFERENCES

Damos, D. L. (1987). *Some considerations in the design of a computerized human information processing battery* (NAMRL Monograph 35). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Delaney, H. D. (1990). Validation of dichotic listening and psychomotor task performance as predictors of primary flight training criteria: Highlighting relevant statistical issues. (NAMRL - 1357) Pensacola, FL: Naval Aerospace Medical Research Laboratory.

SPARTANS

PURPOSE

The SPARTANS (Simple Portable Aviation-Relevant Test-Battery and Answer-scoring System) test battery was developed as a diagnostic battery to assess neurological impairment in high level skills in complex dynamic environments.

DESCRIPTION

The SPARTANS originated in studies reported by Braune and Wickens (1984) and Wickens, Braune, Stokes, and Strayer (1985). The original battery was devised to assess individual differences and age-related performance in pilots and was validated against performance in a motion-based flight simulator.

The SPARTANS can be run on an IBM compatible 286 microcomputer with VGA color graphics and a joystick controller. The current version, 2.0 is modular and each sub-task can be selected from a menu of 20 sub-tasks in any order and can be presented as a practice run or as a data collection run (Banich, Stokes, and Karol, 1990). There are two forms of the battery to permit retesting of subjects. The tests were selected to incorporate those aviation-relevant skills that have been found sensitive to neuropsychological functioning. Spatial ability, for example, was found to be sensitive to dementia of various etiologies (Banich, Stokes, and Elledge, 1988) and is especially disrupted by right hemisphere damage (Perecman, 1983).

Description of the sub-tasks are as follows:

Spatial Memory Task

A set of nonsense or abstract figures designed to ease verbal encoding is presented to the subject (spatial memory task #1). About twenty minutes after the initial presentation, subjects view another set of 40 figures (spatial memory task #2) and must decide if any have been viewed previously.

Hidden figures Recognition Task

This task assesses flexibility of closure. Subjects must decide if an abstract line drawing is embedded within a complex pattern of lines that changes on each trial. The task is always presented in the same orientation. The location of the target pattern on the screen changes from trial to trial during presentations to avoid a decrease in difficulty.

Rotated Hidden Figures Recognition Task

This is a rotated version included to add additional discrimination power in some categories of impairment. Since both a rotated and non-rotated version of the task is presented, the specific effect that rotation has on performance can be determined.

Maze Tracing Task

Subjects are presented with a series of line mazes of increasing complexity. Each maze must be mentally traced to decide if there is an unbroken path from beginning to end. A response is recorded by pressing a "yes" or "no" button on the joystick and both accuracy and latency data are recorded.

Scheduling Task

This task was designed to assess flexibility of strategy and the ability to set priorities. It is similar to the MICROPAT task (Bartram and Dale, 1990). In the task lines "grow" down the screen in any of 5 columns toward a set of boxes containing numbers, whose value is added to the subject's score when reached by a line. The subject decides in which column a line will start by pressing a key from 1 to 5. Several variables fluctuate concurrently - line speed, duration of the box on screen ("box lifetime"), box position, and box value. Subjects must continually make and change decisions to maximize their score. Data is logged on total score and on decision reversals, a variable that is sensitive to cognitive impairment due to stress (Stokes and Raby, 1989).

Spatial Memory Task #2

This is a repeat of Spatial Memory Task #1.

Visual Scanning Task

In this task, subjects estimate the number of Xs printed on the longest of twelve horizontal lines that are presented for a 4 second duration. An average deviation score from the true number of targets is computed over 20 trials.

Sternberg Task

The Sternberg Task is a standard memory search task in which target letters are memorize and a series of stimulus letters are presented. Subjects depress a joystick button to indicate if a stimulus letter has been previously presented in a target set.

Zero Order Pursuit Tracking Task

The zero order (position) tracking task is similar to the First-Order (velocity) Tracking Task. This task is designed to be more sensitive to dysfunction and less influenced by experience than First-Order Tracking.

Dual Task Zero Order Tracking

In this task the zero order tracking task is presented in combination with the Sternberg task to permit dual-task decrements to be measured in both tasks.

First-Order Tracking Task

Subjects are instructed to keep a vertical sinusoidal line between two horizontally movable cursers using the joystick. This task is an adaptation of that developed in the Netherlands by Boer and Gaillard (1986). The task is performed twice, first in a single-task condition and again in a dual-task condition.

Dual Task First-Order Tracking

This task combines the Sternberg task with the first-order tracking task and measures performance decrements due to dual-task timesharing. Sternberg probes are presented randomly to the left or right of the tracking task.

Box Task

This is a measure of risk-taking. Subjects may select up to ten boxes containing hidden scores. The score, once revealed, is added to the subject's box-task total. The subject is aware that one random box per trial is "booby-trapped." If the disaster box is selected, the subject's entire score for that trial is lost. First, the subject's initial performance on trial 1 is recorded before the subject has been influenced by the random pattern of successes and failures across all trials. The mean and standard deviation of the number of boxes selected over 20 trials is calculated. Risk-taking consistency has proved to be sensitive to alcohol induced impairments (Stokes et al., 1990).

Number Subtraction Task

This is a time-stressed attention task adapted from the paper-and-pencil in the Illinois Screening Test. In the present version, a 3-digit number, and then six 3-digit numbers are displayed. Only one number is the correct subtracted value - the others represent incorrect subtractions. The subject selects one value using the joystick, and depresses the joystick button. Subjects proceed at their own pace. Two variables are measured; the total number of subtractions correctly identified in 30 seconds, and the mean response latency.

Digit Symbol Task

This task is also adapted from the Illinois Screening Test. In the present version, symbols (rather than digits) are presented and the subject keys in the associated number. A practice test is given to ensure the instructions are understood. The subject's score is the number of squares correctly filled.

Stroop Test

This is a classic test that measures processing flexibility, conceptualized as the ability to disregard irrelevant information (Stroop, 1935). Subjects are required to identify the color that a particular word spells e. g., red, blue, etc. The test is conducted in two parts. In the control condition, the color words are displayed in black and the subject must match the word with a color patch at the bottom of the screen. In the experimental condition, the terms are displayed

in randomly selected, primary colors that conflict w.:h the color the word spells (e.g., the word red displayed in yellow). The subject must override a powerful perceptual effect to match correctly and rapidly the word to the appropriate color patch. The amount of interference is measured by comparing the time difference and the percent of correct responses in the control and the experimental condition.

Logical Reasoning Task

This task is modified from standard verbal reasoning tests and involves deciding if nonsense syllogisms are valid or invalid. The test consists of 20 nonsense syllogisms, in which both speed and accuracy of response are monitored.

Story Task

This test can be described as the verbal equivalent of the Spatial Memory task. The present form involves the presentation of a short narrative containing several characters and events. The subject then responds to multiple choice questions about the passage. The percentage of correct responses is recorded.

Visual Number Span Task

This task measures working memory capacity, a factor that has been shown to discriminate between good and poor pilot decision making (Wickens et al., 1987. It is adversely affected in individuals with a variety of clinical conditions, including dementia (Albert, Butters, and Brandt, 1981), substance abuse (Nixon, Kujawski, Parsons, and Yohman, 1987), and neurological damage (Lezak, 1979).

Subjects are presented with a set of digits at 1 second intervals accompanied by a brief tone. The length of the digit series increases from one to nine digits as the number of trials progresses. After the screen is cleared the subjects key the digits they recall back into the computer. Correct/incorrect responses are recorded along with two latency variables of initial and final keyboard input.

Reverse Order visual Number Span Task

This task is similar digits forward and reverse of the WAIS-R. Digit span forward is a left hemisphere task, number span in reverse is a right hemisphere task.

PUBLISHER/AGENCY

AMRL, Human Engineering Division
Wright-Patterson AFB. OH
Administration

TIME

Not available

APPARATUS

Not available

PROCEDURE

Not available

RELIABILITY

Not available

CRITERIA

The battery has been used to detect performance decrements under stress.

VALIDITY

Stokes, Belger and Zhang, (1990) found that the SPARTANS in combination with MIDIS, a microcomputer-based flight-decision simulator, was able to show significant decrements in performance under stress for both experienced and inexperienced pilots in non-domain-specific information processing tasks. However, this impairment in information processing was not associated with any performance decrement in simulated flight decision making by experienced pilots.

Banich, Stokes and Karol (1990) concluded that the SPARTANS effectively detects neuropsychological dysfunction and not pilot expertise.

SPARTANS tests were found to be sensitive to the effects of alcohol upon cognitive performance (Stokes, Belger, Banich, and Taylor, 1990), to noise and workload related stress (Stokes and Raby, 1989), also to neuropsychological impairment (Stokes et al., 1988).

Stokes, Banich, and Elledge (1991) concluded that the SPARTANS was not an appropriate instrument for large scale screening of flight medical certificate applicants. This was based on their discriminate validity study comparing pilots and subjects in a clinical population. They did suggest that the SPARTANS might be useful when combined with traditional neuropsychological testing when examining someone suspected of impairment.

NORMS

Data may be processed using a utility program that provides for graphic display of normative data.

SCORING

SUMMARY

The SPARTANS consists of 20 subtests, (1) Spatial Memory Task (2) Hidden Figures Task, (3) Rotated Hidden Figures Task (4) Maze Tracing Task, (5) Schedule Task (6) Spatial Memory Task #2, (7) Visual Scanning Task, (8) Sternberg Task, (9) Tracking Task (zero order), (10) Dual Task (Sternberg with zero order tracking), (11) Tracking Task (first order), (12) Dual Task (Sternberg with first order tracking), (13) Box (risk-taking) Task, (14) Subtraction Task, (15) Digit-symbol Task, (16) Stroop Task, (17) Syllogisms Task, (18) Story Task, (19) Visual Number Span Task and, (20) Backwards Visual Number Span Task (Banich, Stokes, and Karol, 1990.)

REFERENCES

Albert, M. N., Butters, N., and Brandt, J. (1981). Development of remote memory loss in patients with Huntington's disease. Journal of Clinical Neuropsychology, 1-12.

Banich, M. T., Stokes, A. F., & Elledge, V. (1987). Cognitive function evaluation in the medical certification of airmen--a literature review. (Final Tech. Rep. ARL-88-3/FAA-88-1) Savoy, Il: University of Illinois, Aviation Research Laboratory.

Banich, M. T., Stokes, A. F., and Elledge, V. C. (1988). Neuropsychological screening of aviators: A review. Aviation, Space and Environmental Medicine, 60, 361-366.

Banich, M. T., Stokes, A. F., & Karol, D. (1990). Cognitive function evaluation in the medical certification of airmen phase B. (ARL-90-2/FAA-90-1), Oklahoma City: FAA Aeromedical Research Branch.

Bartram, D., and Dale, H. C. (1983). Micropat version 3: A description of the fully automated personnel selection testing system being developed for the Army Air Corps. (Report ERG/Y6536/83/7). Ergonomics Research Group, University of Hull.

Boer, L. C., and Gaillard, A. W. K. (1986). TASKOMAT: A standardized task battery. Soesterberg, Netherlands: Institute for Perception TNO.

Braune, R., and Wickens, C. D. (1984). Individual differences and age related performance assessment in aviators. Part II: Initial validation (Tech. Rep. NAMRL-84-1/EPL-84-1). Savoy, IL: University of Illinois Engineering Psychology Laboratory.

Lezak, M. D. (1979). Recovery of memory and learning functions following traumatic brain injury. Cortex, 15, 63-70.

Nixon, S. J., Kujawski, A., Parsons, O. A., and Yohman, J. R. (1987). Semantic (verbal) and figural memory impairment in alcoholics. Journal of Clinical and Experimental Neuropsychology, 9(4), 311-322.

North, R. A., and Gopher, D. (1976). Measures of attention as predictors of flight performance. Human Factors, 18(1), 1-14.

Perecman, E. (1983). Cognitive functioning in the right hemisphere. New York: Academic Press.

Poon, L. W. (Ed.). (1980). Aging in the 1980's. Washington, DC: American Psychological Association.

Salthouse, T. A. (1982). Adult cognition: An experimental psychology of human aging. New York: Springer-Verlag.

Salthouse, T. A., and Somberg, B. L. (1982). Isolating the age deficit of speeded performance. Journal of Gerontology, 37, 59-63.

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. F., Banich, M. T., & Elledge, V. (1988). Cognitive function evaluation in the medical certification of airmen. (Final Tech. Rep. ARL-88-4/FAA-88-2) Savoy, Il: University of Illinois, Aviation Research Laboratory.

Stokes, A. F., Banich, M. T., Elledge, V. C., & Ke, Y. (1988). Cognitive function evaluation in the medical certification of airmen (Final Tech. Rep. ARL-88-4/FAA-88-2). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Stokes, A. F., Belger, A., & Zhang, K. (1990). Investigation of factors compromising a model of pilot decision making: Part II. Anxiety and cognitive strategies in expert and novice aviators. (Final Tech. Rep. ARL-90-8/SCEEE-90-2) Wright-Patterson AFB: AMRL, Human Engineering Division.

Stokes, A. F., & Raby, M. (1989). Stress and cognitive performance in trainee pilots. Proceedings of the 33rd Annual Meeting of the Human Factors Society (pp. 883-887). Santa Monica, CA: Human Factors Society.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Wickens, C. D., Braune, R., Stokes, A., and Strayer, D. (1985). Individual differences and age related change: Refinement and elaboration of an information processing performance battery with aviation relevant task structure (Final Rep. EPL-85-1/NAMRL-85-1). Champaign, IL: University of Illinois Engineering Psychology Laboratory, Department of Psychology.

UNIFIED TRI-SERVICES
COGNITIVE PERFORMANCE ASSESSMENT BATTERY

PURPOSE

The Unified Tri-Services Cognitive Performance Assessment Battery (UTC-PAB) was developed to access cognitive performance in a multiple level drug evaluation program. It also may have potential for the assessment of neuropsychological deficits, stress effects and pilot selection (Perez, Masculine, Ramsey, and Urban, 1987).

DESCRIPTION

The UTC-PAB was drawn from several other batteries, primarily from components of the Army (WRAIR-PAB), Navy (PETER, NAVAL CCT) or Air Force (CTS) batteries. The 25 sub-tests are as follows:

LINGUISTIC PROCESSING TASK (VISUAL AND SEMANTIC CODING)

PURPOSE

The Linguistic Processing Task was designed to measure the ability to code linguistic information at different depths of process and transform linguistic information.

DESCRIPTION

This task is a based on Posner and Mitchell's (1967) letter match task and depth of processing tasks (e.g., Craik and Tulving, 1975). It is a standardized loading task requiring the classification of pairs of letters or words. Letter or word pairs are presented on a monitor, with instructions to respond "same" if the items match or "different" if they do not. Three levels of task demand are presented: (1) Physical letter match, requiring letter pairs to be physically identical to match (low demand); (2) category match, requiring that both letters to be either consonants or vowels (moderate demand); and (3) antonym match, requiring only words opposite in meaning to form a match (high demand). Each set lasts three minutes.

GRAMMATICAL REASONING (TRADITIONAL) (LOGICAL REASONING)

PURPOSE

The grammatical reasoning test was designed to measure general reasoning ability. It is a type of sentence verification task that tests the processing capacity of working memory. It is also sensitive to environmental stress, pollutants, and the effects of sleep loss.

DESCRIPTION

First, pairs of letters (AB or BA) and a statement about their sequential arrangement are presented to the subject. Then the task is to decide if the statement and letter pairs match or fail to match. Subjects respond by pressing one of two buttons on a keypad that are labeled TRUE and FALSE.

GRAMMATICAL REASONING (SYMBOLIC) (LOGICAL REASONING)

PURPOSE

This task was designed to measure general reasoning ability. The symbolic grammatical reasoning task is a type of sentence verification task that taps the processing capacity of working memory. It is sensitive to variable information processing demands and is probably sensitive to environmental stress, pollutants, and sleep loss.

DESCRIPTION

The symbolic grammatical reasoning task tests variable demands required for the manipulation and comparison of grammatical information. It is based on Baddeley's (1968) Grammatical Reasoning Task. The stimuli are sentences of varying syntactic structure simultaneously presented with two or three sets of symbols (e.g., *, @, and #). The sentences must be analyzed to find if they correctly describe the ordering of the characters in the symbol set. The amount and complexity of grammatical analysis decide task demand. Three different levels of task demand are measured, they are: (1) single-sentence items of variable syntactic construction that describe the order of symbol pairs (i.e., all possible stimuli from the Baddeley version, and substituting symbols for letters) -- low demand; (2) two-sentence items worded actively and positively, that describe the relative positions of three symbols--moderate demand; and (3) two-sentence items worded either actively/negatively or passively/negatively that describe three symbols--high demand.

TWO-COLUMN ADDITION (NUMBER FACILITY)

PURPOSE

This subject-paced, mental arithmetic test was designed to measure the ability to perform simple addition. It measures the ability to quickly and accurately retrieve arithmetic information and to use procedural knowledge. It also measures short term memory.

DESCRIPTION

First, a set of 45 trials is presented to the subject on the monitor. Each trial consists of three 2-digit numbers simultaneously presented on a monitor in a column format. Subjects are required to sum and enter their responses as rapidly as possible on a keyboard. Responses are entered beginning with the left hand digit first (usually the hundreds and tens digit). The column

of digits disappears with the first valid key entry; requiring the subject to know the entire answer before entering a response. A trial is ended by pressing the return key or when a deadline period of 15 seconds has elapsed. Subjects receive speed/accuracy feedback during the training trials but do not receive feedback during the experimental trials.

MATHEMATICAL PROCESSING TASK (NUMBER FACILITY/GENERAL REASONING)

PURPOSE

This self-placed mental arithmetic task was designed to test a subject's information processing resources associated with working memory. The subject must: (a) retrieve information from long term memory, (b) update information in working memory, (c) sequentially execute different arithmetic operations, and (d) perform numeric comparisons.

DESCRIPTION

Subjects perform one or more addition and/or subtraction operations on single digit numbers and decide if the answer is greater or less than five. Problems are presented in the center of the screen in a horizontal format for left to right solution and are followed by an equal sign. The responses are entered via a two button keypad.

Three versions of this task can be selected and are designed to produce different response time performance. Each version requires three minutes of continuous performance and reaction times are recorded. There are three versions of this test: (1) low demand version--problems containing only one mathematical operation, (2) moderate demand version--problems containing two mathematical operations, and (3) high demand version--problems containing three mathematical operations.

CONTINUOUS RECOGNITION TASK (WORKING MEMORY-ENCODING and RECOGNITION)

PURPOSE

The Continuous Recognition Task was designed to tap variable demands associated with encoding and storage in working memory. The task tests the ability to encode, rehearse, recall, and compare numbers in short term memory on a continuous basis.

DESCRIPTION

The memory test consists of a random series of numbers that must be encoded in a sequential fashion. As each number in the series is presented, a probe number is presented simultaneously. The operator must compare this probe number to a previous item at a specified position back in the series. The subject must decide if the previously presented item is the same as or different from the probe number. Thus, the task taps working memory by requiring

subjects to maintain accurately, update, and access a store of information on a continuous basis. Task difficulty is manipulated by varying the number of digits comprising each item, and the length of the series that must remembered to respond to recall probes.

FOUR-CHOICE SERIAL REACTION TIME (ENCODING, CATEGORIZATION, RESPONSE SELECTION)

PURPOSE

This task was designed to measure stimulus encoding and categorization, and response selection. It probably tests encoding more heavily.

DESCRIPTION

A blinking plus sign imposed on the cursor is presented in one quadrant of a monitor. The subject presses a key on the keyboard that corresponds to the quadrant with the blinking "+." The blinking "+" remains in a quadrant until the correct key is pressed and then randomly reappears in another quadrant. If no button is pressed within 2.5 seconds, a bell rings at 0.1 second intervals until a response is made. Subjects are instructed to respond as quickly and accurately as possible. The task lasts 6 minutes.

ALPHA-NUMERIC VISUAL VIGILANCE TASK (SUSTAINED VISUAL ATTENTION--CHOICE RT)

PURPOSE

The Alpha-Numeric Visual Vigilance Task (ANVVT) was designed to test the ability to make rapid decisions in response to visual symbols for long nonstop periods. The ANVVT is a discrimination reaction task intended as a simulation of a person monitoring a visual display under conditions that might produce fatigue and performance decrement without the subject being aware of it.

DESCRIPTION

In this version, random alphabetic characters or numbers are presented at random intervals on a monitor ranging between 6 and 14 seconds, with an average interval of 10 seconds. The number or character is 10 by 28 mm in size and remains on the screen for 500 msec. Subjects press a hand held push button with their thumb every time an "A" or a "3" appears. No other responses are
required.

Twenty "As" and "3s" are randomly inserted within 160 other characters and numbers presented during this 30-minute task. Response latencies and errors are recorded. There are two types of possible errors: (1) errors of commission (responding to non "As" and non "3s"), and (2) errors of omission (not responding to an "A" r "3" in 5 seconds). Reaction times are

recorded in msec for correct responses and errors of commission. Errors of omission are cored as reaction times of 5000 msec.

MEMORY SEARCH TASKS (SHORT TERM WORKING MEMORY--AUDITORY AND VISUAL MODALITIES)

PURPOSE

This memory search task was designed to test the ability to make comparisons of letters stored in memory. The task diagnoses selective retrieval and comparison in short term working memory. It also may tap the encoding of stimulus items, categorization, response selection, and response execution.

DESCRIPTION

A "positive set" of either one, two, four, or six alphabetic characters is presented to be stored in memory. The remaining alphabetic characters make up the "negative set." Following the presentation of the "positive set," individual probe letters are presented for comparison and classification as members of the positive set or the negative set. Subjects respond by pressing a key on a two button keypad.

There are three different versions of this task. Each version is presented in a visual format and an auditory format for a total of six unique versions. A different positive set is generated on every trial in the varied set procedure (VS) followed by a single probe item. In the fixed set procedure (FS) the positive set is followed by 100 probes to form a trial. A trial in the mixed set procedure (MS) consists of 10 separate positive sets of equivalent size, a&h followed by 10 probes for correct classification. In the visual format all stimuli are presented on a monitor. In the auditory versions the probe items are presented with a speech synthesis system and positive sets are presented on a monitor and the speech synthesis system.

SPATIAL PROCESSING TASK (SPATIAL ORIENTATION/ROTATION SHORT TERM MEMORY)

PURPOSE

This task was designed to measure the ability to rotate mentally a series of histograms before making a same/different judgement. Since the standard and test stimuli are presented successively rather than simultaneously the test is believed to tap visual short term memory.

DESCRIPTION

A series of histograms is presented one at a time. The subject must decide if the second histogram of each pair is identical with the first. The subject answers by either pressing a button labeled "same" or "different" on a two key response box. Task loadings are varied by presenting a two bar standard stimulus with the test stimulus in the zero degree orientation for

low loading; a four bar standard with the test stimulus in the 90 or 270 degree orientation for a moderate task loading; and a six bar standard with the test stimulus in the 180 degree orientation for a high task loading.

MATRIX ROTATION TASK (SPATIAL ROTATION SHORT TERM MEMORY)

PURPOSE

The Matrix Rotation Task was designed to assess spatial rotation ability. Spatial rotation or spatial transformation is one component of spatial orientation. This task also taps short term perceptual memory.

DESCRIPTION

A series of 5 by 5 cell matrices, are presented at the center of the display. Each matrix has five illuminated cells. After a pause, the screen blanks and a second matrix is presented. The subject must decide if the second matrix is identical with the first. Responses are entered on a two key response box.

A matrix is considered identical only if it is a 90 degree rotation of the previous matrix. Successive test matrices are always presented in a different orientation.

MANIKIN TEST (SPATIAL ORIENTATION ROTATION ABILITY)

PURPOSE

The Manikin Test was designed to assess the ability to rotate a mental image and related transformations. This ability is one measure of three general subdivisions of spatial ability. Lohman (1979) called this ability spatial orientation (SO). It requires mental movement of the self to view a test stimulus from a different perspective.

DESCRIPTION

The Manikin Test consists of a series of 64 trials. On each trial, a human figure (the manikin) is displayed on a monitor. The figure will be in four orientations: (1) facing toward the subject; (2) facing away from the subject; (3) right side up; or (4) upside down. Combinations of all possible pairs of these positions yield a block of 16 possible orientations.

The manikin holds a red or blue box in each hand. The manikin stands on a platform that matches the color of a box in his hand. The task is to decide which hand is holding the box that matches the platform color. Responses are entered on a box with two labeled buttons, one "left hand" and one "right hand." During the 64 training trials (four presentations of each orientation) the subject receives feedback, but no feedback is given during the test trials.

PATTERN COMPARISON (SIMULTANEOUS) (PERCEPTUAL SPEED PATTERN COMPARISON)

PURPOSE

The self-paced pattern comparison test was designed to assess perceptual speed. Perceptual speed is one aspect of general spatial ability. The test taps the ability to make simultaneous judgements about the similarity of two patterns.

DESCRIPTION

There are 60 trials in the test. On each trial, a monitor displays two patterns of eight dots, side by side. The pattern on the left is the standard. The task is to decide if the pattern on the right is identical with the standard. A response on the response box terminates the trial. If no response occurs before the end of a 15-second deadline, the trial is terminated automatically. Speed and accuracy feedback is given to the subjects during the 10 training trails. No feedback is given during the test trials.

PATTERN COMPARISON (SUCCESSIVE) (PERCEPTUAL SPEED SHORT TERM SPATIAL MEMORY)

PURPOSE

This task was designed to measure the short term spatial memory and perceptual speed. The test diagnoses spatial memory, since the subject must store the standard in memory and compare it with the test pattern.

DESCRIPTION

The test consists of a series of 60 trials. Each trial proceeds as follows: The standard pattern is presented for 1.5 seconds. At the end of each period, the screen clears for 3.5 seconds and the test pattern appears. The test pattern remains on the display until the 15 second deadline or if the subject makes a response. The task is to decide if the two dot patterns are the same or different.

During the training phase, the subject responds to 10 trials. Response speed and accuracy feedback are provided to the subject after each training trials. Feedback is not given during the test trials.

VISUAL SCANNING TASK (PERCEPTUAL SPEED)

PURPOSE

This task is based on Neisser's (1963) letter search task that required subjects to search for and detect a target embedded in non target items. This test was designed to diagnose the ability to perform rapid visual pattern discrimination.

DESCRIPTION

The UTC-PAB visual scanning task has two versions. Both versions require the subject to scan visually a matrix of letters (25 rows by 5 columns) in normal reading order (left to right, top to bottom) to detect a target letter (e.g., "K") embedded in the matrix. In the lite pen version, when the target letter is detected, the subject identifies the exact location of the target using a lit pen. In the keyboard version, the subject identifies the row of the matrix where the target is embedded with a keypad.

CODE SUBSTITUTION TASK (PERCEPTUAL SPEED, ASSOCIATIVE LEARNING ABILITY)

PURPOSE

This task was designed to tap the ability to encode rapidly and evaluate stimuli.

DESCRIPTION

The UTC-PAB code substitution task is based on a paper and pencil version of the task contained within the Wechsler Adult Intelligence Scale (Wechsler, 1958), and was designed to assess associative learning ability and perceptual speed. A string of nine letters and a string of nine digits is presented on a monitor with the digit string positioned immediately below the letter string. Each digit corresponds to a given letter. A test letter is then presented at the bottom of the screen, below the two coding strings. The subject has to decide which digit corresponds to that test letter in the coding strings by pressing a key on a numbered keypad. The letter-digit associative pairings are the same for the entire test.

VISUAL PROBABILITY MONITORING TASK (SPATIAL SCANNING/SIGNAL DETECTION)

PURPOSE

This task was designed to measure the ability to scan and detect visual signals.

DESCRIPTION

A display of dials is presented to the subjects with instructions to monitor the movement of a pointer located below each dial. Under normal conditions, the pointer moves randomly from one position to another to simulate the pointer fluctuations on an actual dial. At unpredictable intervals, the pointer begins to move non-randomly, staying predominantly to the left or right half the dial. These biases in pointer movement are the targets or "signals" that the subject is to respond. The task is to detect the presence of a "signal" and press the correct key after the biased dial has returned to the original random pointer movement.

The test has three task demand levels based on the number of dials that are displayed and the discriminability of the signals. A test trial consists of 3 minutes of continuous monitoring and only one signal can be present at a time. Signals may occur any time within a trial with the restriction of a 25 second minimum separation of the offset of a signal and the onset of the following signal. Test trials usually contain two or three signals. In conditions where three or four dials are monitored (Moderate Task Level and High Task Level) the signal display dial is randomly selected.

TIME WALL (TIME ESTIMATION) TASK

PURPOSE

The time wall task was designed to measure the ability to estimate the time a moving target, moving at a constant rate, will travel a predetermined distance. On each trial the subject must integrate the available speed and distance information to anticipate the time the target will reach a certain spot on the screen.

DESCRIPTION

The UTC-PAB time wall task is a nonverbal time estimating task. A small object moving at constant velocity passes behind an opaque barrier and the subject must estimate the time when the object will reappear. The time wall differs from other time estimation tasks because the discrete mediating responses such as counting or tapping are of no direct obvious aid. In this version, movement is vertical not horizontal for purposes of visual field symmetry. The barrier contains a hold or notch the same shape and size as the object, and the subject estimates the moment when the entire notch will be filled. This implementation uses a nominal 10- second interval.

INTERVAL PRODUCTION TASK (RESPONSE TIMING)

PURPOSE

This task was designed to be used as a secondary task to measure demands placed on motor output by a primary task (Michon, 1966). However, it may be used as a stand alone test

to examine The degree to which variables such as drugs, environmental stress, and toxic substances disrupt manual response timing.

DESCRIPTION

The subject is required to generate a series of intervals by tapping a finger key at a rate of one to three responses per second. The subject taps with the forefinger of the preferred hand on a paddle shaped key. The subject must maintain equal intervals by tapping at as regular a rate as possible. Interval response and intervals of less than 10 msec are rejected as spurious input. The task is run in 3-minute trials.

STROOP TEST (INTERFERENCE SUSCEPTIBILITY TO RESPONSE COMPETITION)

PURPOSE

This test is based on the classic color-word test developed by Stroop (1935). It was designed to measure susceptibility to response interference.

DESCRIPTION

Both color and non-color words are presented one at a time on a monitor. All words are displayed in red, blue, or green and the subject is must press a color coded key that corresponds to the color that is presented.

Three versions of this test are available for selection and each is designed to produce different response time performance.

The following is a brief description of the three test versions:

1. The control Version of this test contains three possible stimuli. This version of the test is intended to be used with the Interference Version; however, it may be used by itself as a choice reaction time task.

2. The Interference Version contains six CWI (color-word incongruent) stimuli. This version represents the usual interference condition found in the Stroop color-word test.

.   The Combined Version used six CWI and six NW (neutral words) stimuli. This version of the test represents the accepted procedure that is employed in the examination of response interference. That is, stimuli that are free of response interference (e.g., NW) are presented with those that produce maximum interferences (e.g., CWI). The difference in reaction time between CWI and NW shows response interference where such factor as stimulus encoding and response generation have been equated.

DICHOTIC LISTENING TASK (AUDITORY SELECTIVE ATTENTION)

PURPOSE

This test was designed to evaluate auditory selective attention.

DESCRIPTION

Subjects must attend to a list of letters and digits that is presented to one ear while trying to ignore similar information being presented to the other ear. The stimuli are produced by a computer controlled speech synthesizer and are presented over dual channel headphones.

UNSTABLE TRACKING TASK (MANUAL RESPONSE CONTROL)

PURPOSE

This task was designed to measure rapid and accurate manual responses.

DESCRIPTION

A monitor displays a fixed target area at the center. A cursor moves vertically from this target while the operator attempts to keep the cursor centered over the target via rotary movement of a control knob. Three reliably different demand levels have been established by Shingledecker (1984), based on two tracking performance measures (average absolute tracking error and number of control losses) and a subjective measure (task difficulty ratings).

MEMORY SEARCH-TRACKING COMBINATION (TIME SHARING ABILITY)

PURPOSE

This dual task combination was designed measure time sharing ability; that is, the ability to perform two tasks concurrently.

DESCRIPTION

This is a dual task paradigm involving unstable tracking and the Sternberg Memory Search Task as employed by Wickens and Sandry (1982). Subjects must track with their left hand and respond to the memory search stimuli with their right hand.

At the beginning of each trial, the subject is shown the positive set of the Sternberg task, as under single task conditions. This display is erased and the trial begins 2 seconds later. Subjects are told to respond quickly and accurately and that both tasks are equally important.

MATCHING TO SAMPLE (SPATIAL MEMORY PATTERN RECOGNITION)

PURPOSE

This task was designed to assess the ability to choose quickly and accurately a test stimulus that is identical with a standard stimulus. The test measures short term spatial memory and pattern recognition skills.

DESCRIPTION

The subject is shown a single 4 by 4 matrix centered on the monitor. The matrix has either red or yellow cells. The number of cells of each color is randomly determined for each stimulus. After viewing the sample stimulus for a time adequate for committing the stimulus to memory, the subject begins the test trial. The test trial consists of two 4 by 4 matrices, side by side on the screen. One of the matrices is identical with the previously presented standard stimulus, while the other is different. The task is to select the test stimulus that matches the standard. There are 30 trials.

ITEM-ORDER TEST (SHORT TERM MEMORY RECOGNITION)

PURPOSE

The item-order test was designed to examine the ability to recognize strings of letters as either the same or different. Error rates produced from this test should reflect short term memory recognition.

DESCRIPTION

In the item-order test, the subject is presented a target string of 7 consonants on a monitor. The target string is displayed for 2 seconds and then the screen goes blank for 2.5 seconds. Immediately following the blank display, a new string of letters is presented. The second letter string is the test string. The subject must decide if the test string is identical with the target string. Subjects make their responses by pressing one of two buttons, labeled "same" or "different." The test string has three possible relationships to the target string: (1) the two strings are identical, (2) the same letters are in the two strings but the letters are in a different order, or (3) the two strings have different letters. A single target string-test string pair forms one trial. The test has 40 trials. The dependent variables are response accuracy and response latency.

PUBLISHER/AGENCY

Armstrong Aeromedical Research Laboratory
Human Systems Division
Wright-Patterson AFB, OH 45433-6573

ADMINISTRATION

Time

The time required would depend on the number of sub-tests used.

Apparatus

MS-DOS, IBM-PC compatible microcomputer.

Procedure

This would vary with the number of sub-tests used.

RELIABILITY

The reliability of the complete battery is not available.

CRITERIA

Not available

VALIDITY

Not available

NORMS

Not available

SCORING

Not available

SUMMARY

The sub-tests were selected based on two dimensions thought to be critical in the effects of drugs on cognitive performance: (1) the stage of processing most markedly affected by task demands; and (2) the requirement to attend selectively to different sources of information.

There are six sub-tests in the UTC-PAB designed to measure the desired cognitive functions: (1) Perceptual Input, Detection, and Identification - visual scanning task, visual probability monitoring task, pattern comparison (simultaneous), and four-choice serial reaction time; (2) Central Processing - auditory memory search (memory search tasks), continuous recognition task, code substitution task, visual memory search (memory search tasks), and item order test; (3) Information Integration/Manipulation- Linguistic/Symbolic - linguistic processing task, two-column addition, grammatical reasoning (symbolic), mathematical processing task, and

grammatical reasoning (traditional); (4) Information Integration/Manipulation-Spatial Mode - spatial processing task, matching to sample, time wall, matrix rotation task (spatial processing task), manikin test, and pattern comparison (successive); (5) Output/Response Execution - interval production task, and unstable tracking task; and (6) Selective/Divided Attention-dichotic listening task, memory search/unstable tracking combination (Sternberg-tracking combination), and Stroop test.

- Advantages

-- Minimum hardware requirements

- Disadvantages

-- None noted

REFERENCES

Baddeley, A. D. (1968). A 3-minute reasoning test based on grammatical transformation. Psychonomic Science, 10, 341-342.

Carter, R. C., Kennedy, R. S., & Bittner, A. C. Jr. (1981). Grammatical reasoning: A stable performance yardstick. Human Factors, 23, 587-591.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 04, 268-294.

Hord, D. J. (1982). An EEG predictor of performance decrement in a vigilance task. (Tech. Rep. No. 82-2) San Diego: Naval Health Research Center.

Keele, S. W., & Hawkins, H. L. (1985). Explorations of individual differences relevant to high level skill. Journal of Motor Behavior, 14, 3-23.

Kennedy, R. S., Dunlap, M. B., Jones, N. E., Lane, N. E., & Wilkes, R. L. (1985). Portable human assessment battery: Stability, reliability, factor structure, and correlation with tests of intelligence. Final Tech. Rep., National Science Foundation.

Kennedy, R. S., Wilkes, R. L., Lane, N. E. & Hamick, J. L. (1985). Preliminary evaluation of a micro-based repeated measures testing system. Essex Orlando Tech. Rep. EOSR-85-1.

Lohman, D. F. (1979). Spatial ability: A review and reanalysis of the correlational literature. (Tech. Rep. No. 8), Stanford, CA: Aptitude Research project, School of Education, Stanford University.

Michon, J. A. (1966). Tapping regularity as a measure of perceptual motor load. Ergonomics, 9, 401-412.

Neisser, V. (1963). Decision-time without reaction-time: Experiments in visual scanning. American Journal of Psychology, 76, 376-385.

Perez, W. A., Masculine, P. L., Ramsey, E. G. & Urban, K. E. (1987). Unified Tri-Services Cognitive Performance Assessment Battery: Review and methodology (U). (AAMRL-TR-87-007) Wright-Patterson, AFB, OH: Armstrong Aeromedical Research Laboratory.

Posner, M. I. & Mitchell, R. F. (1967). Chronometric analysis of classifications. Psychological Review, 74, 392-409.

Seales, D. M., Kennedy, R. S., & Bittner, A. C. (1980). Development of performance evaluation tests for environmental research.

Shingledecker, C. A. (1984). A task battery for applied human performance assessment research. (Tech. Rep. No. AFAMRL-TR-84-071), Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratories.

Stroop, J. R. (1935). Studies of interference in verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Wechsler, D. (1958). The measurement and appraisal of adult Intelligence. Baltimore: The Williams and Wilkins Co.

Wickens, C. D. & Sandry, D. (1982). Task-hemispheric integrity in dual task performance. Acta Psychologica, 52, 227-247.

APPENDIX B

Figures Plotting Age, Flight Time, and Accidents

Golaszewski, 1983

1976-1980 Class III Pilots
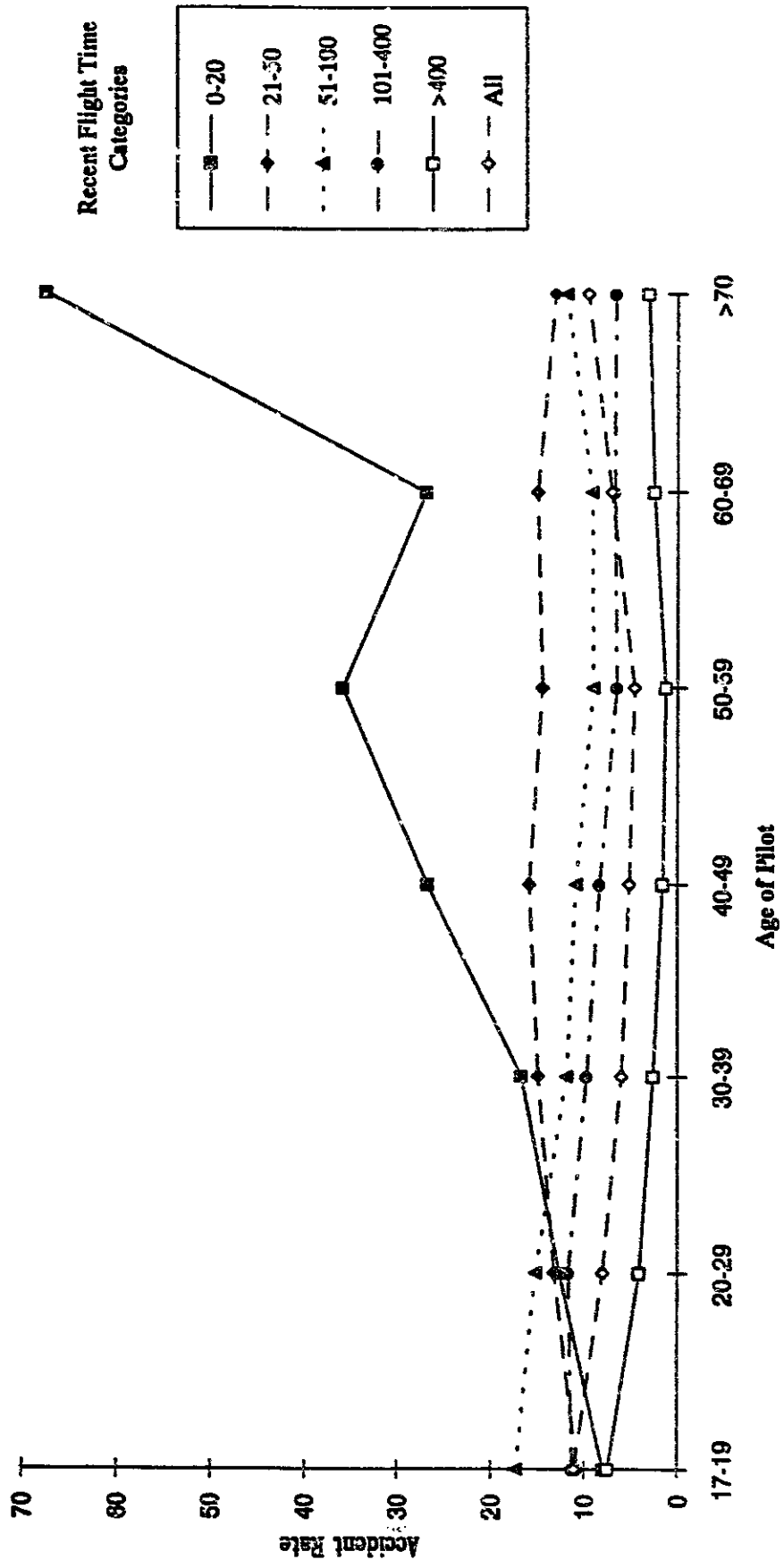## Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 1)
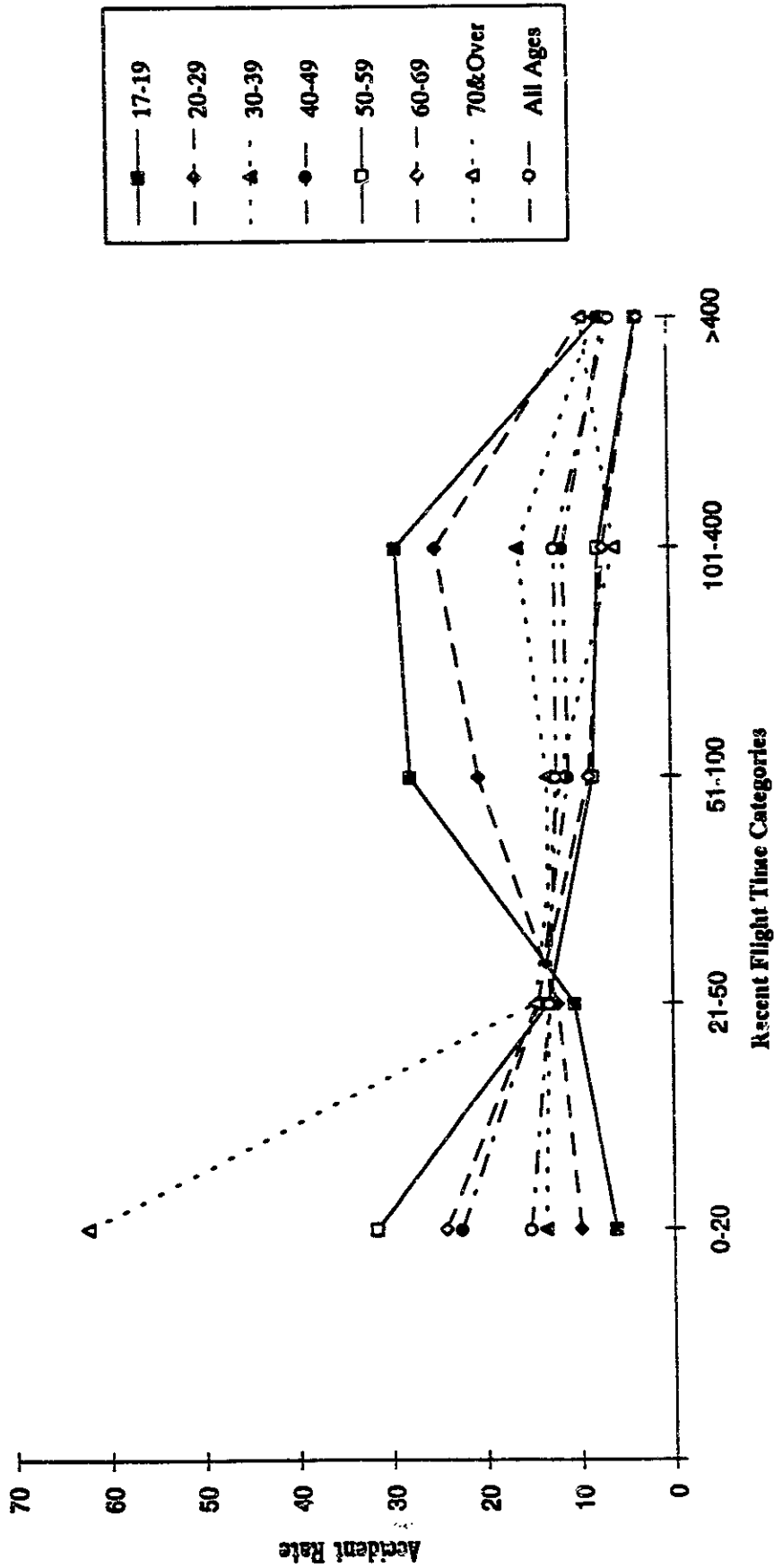
1976-1980 Class I, II, and III Pilots
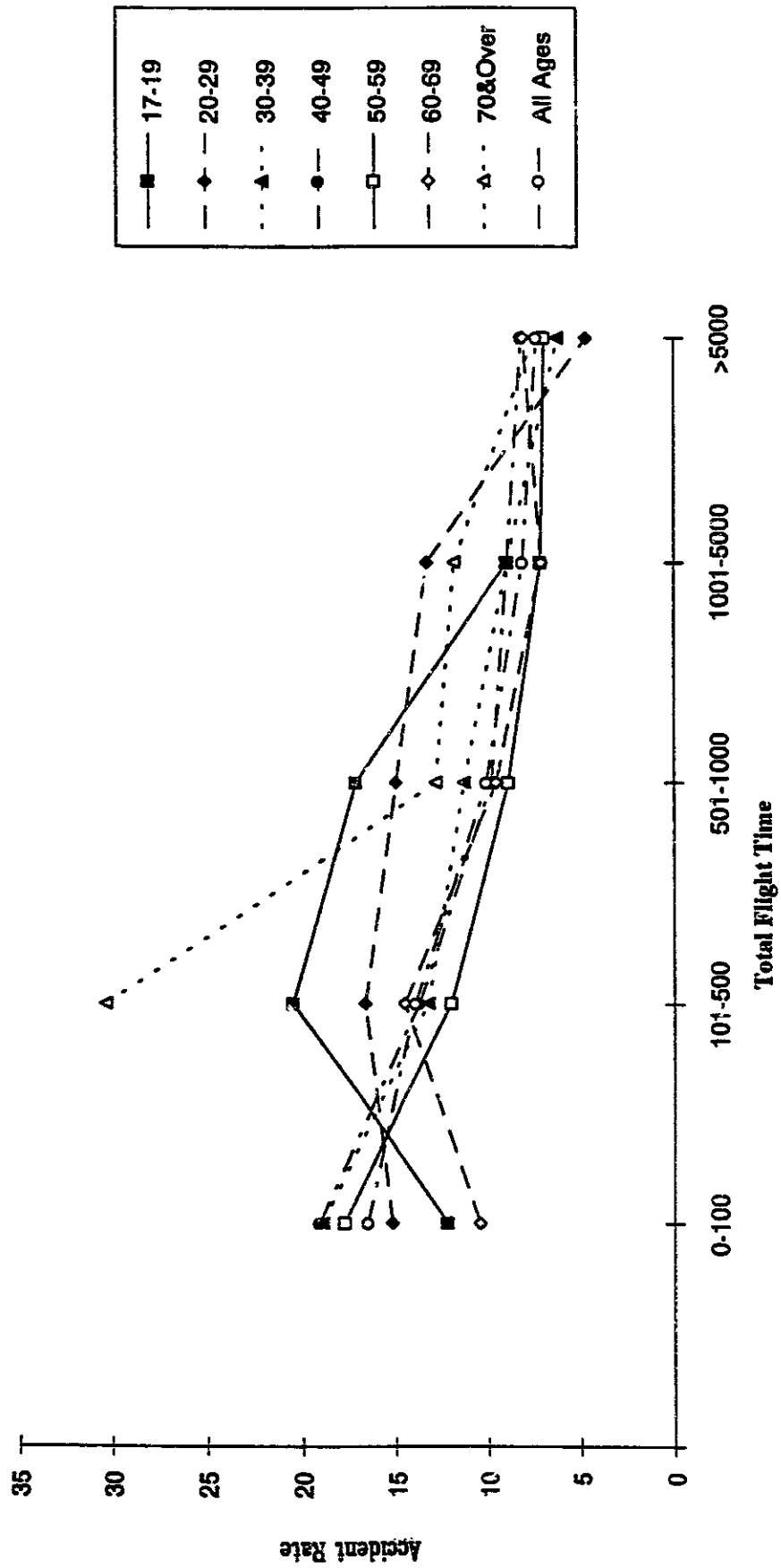Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 2)
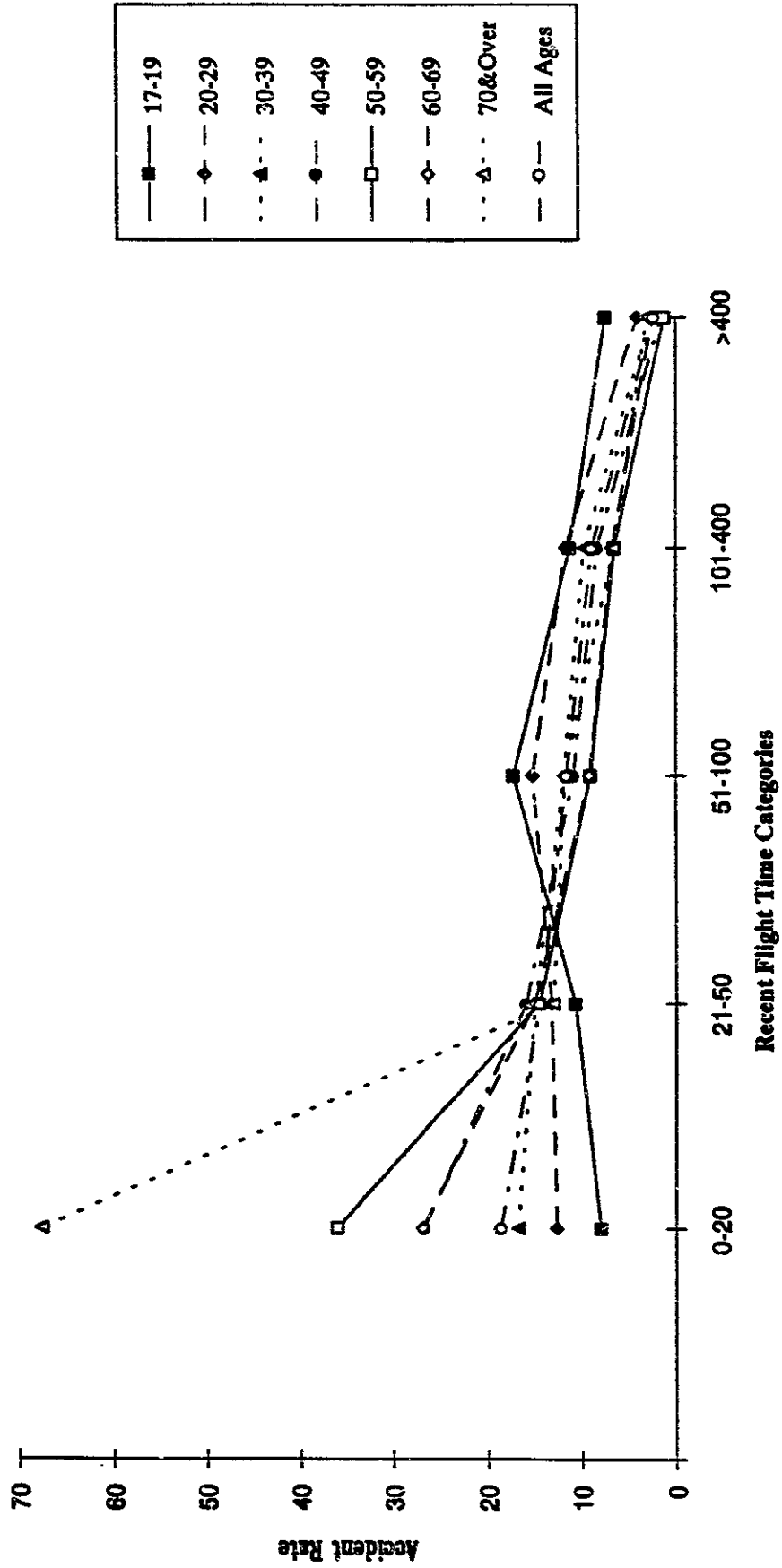
1976-1980 Class III Pilots
Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 3)
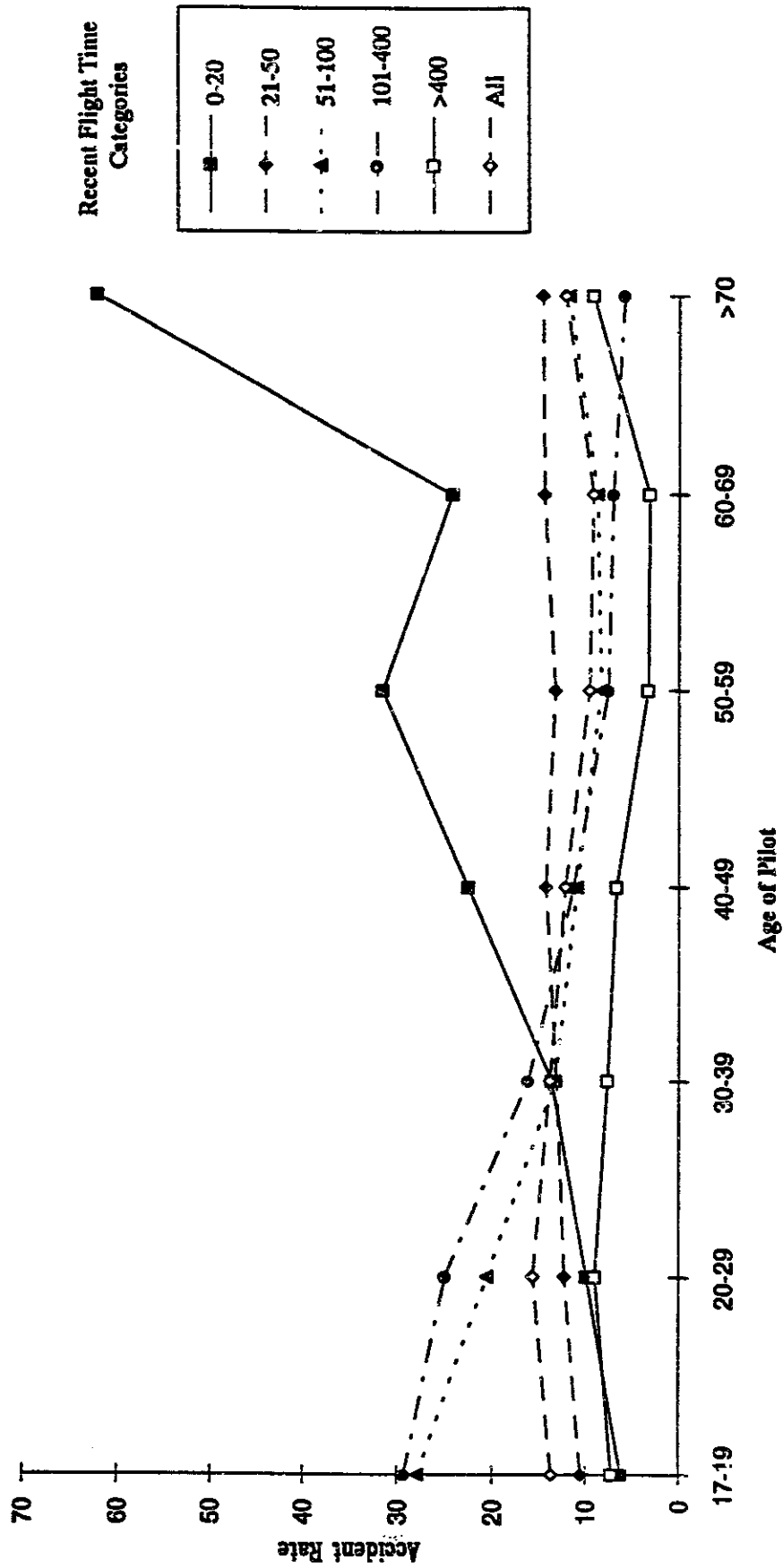
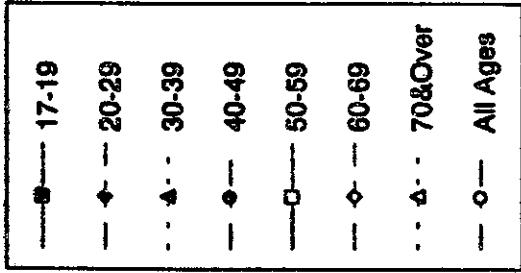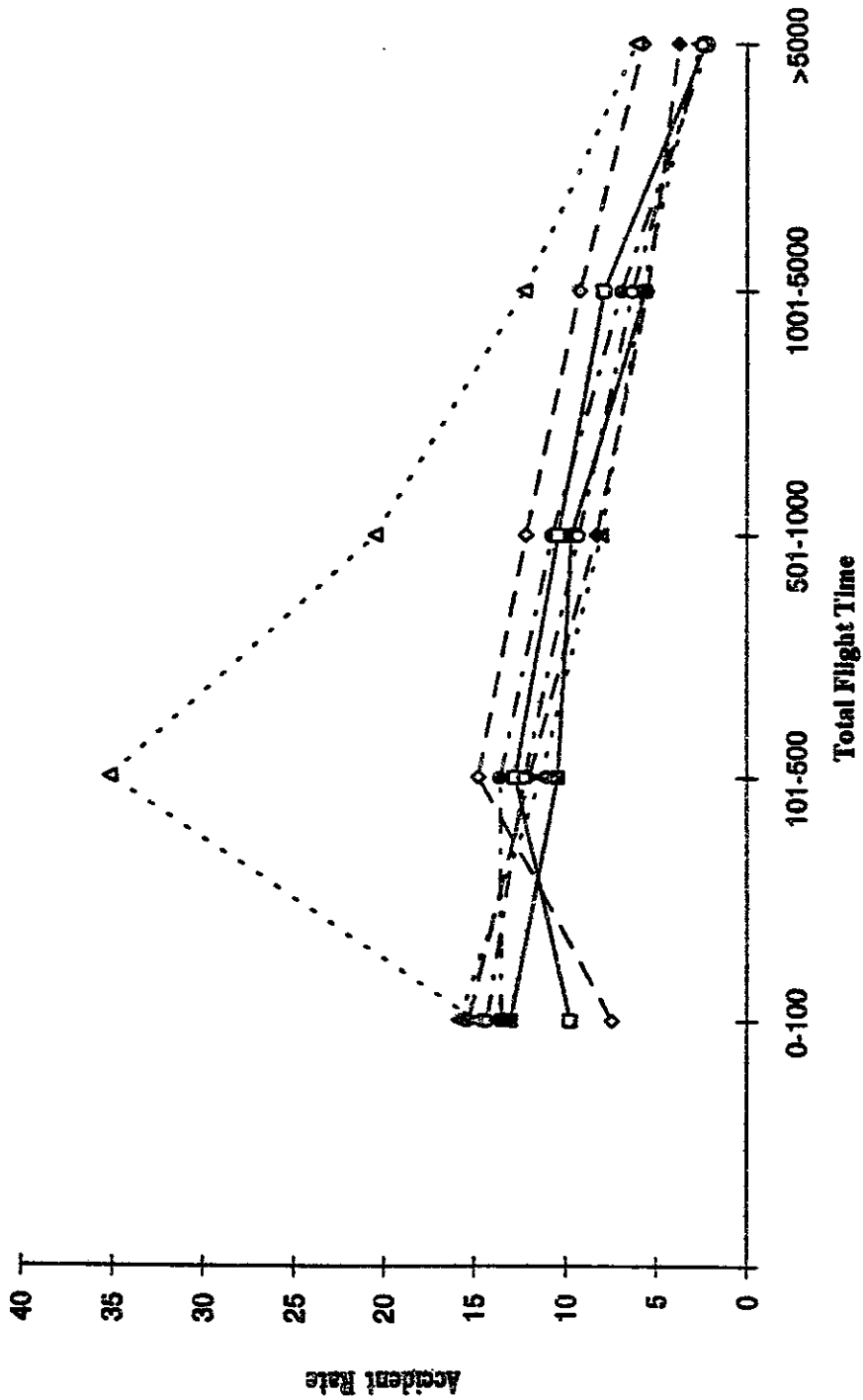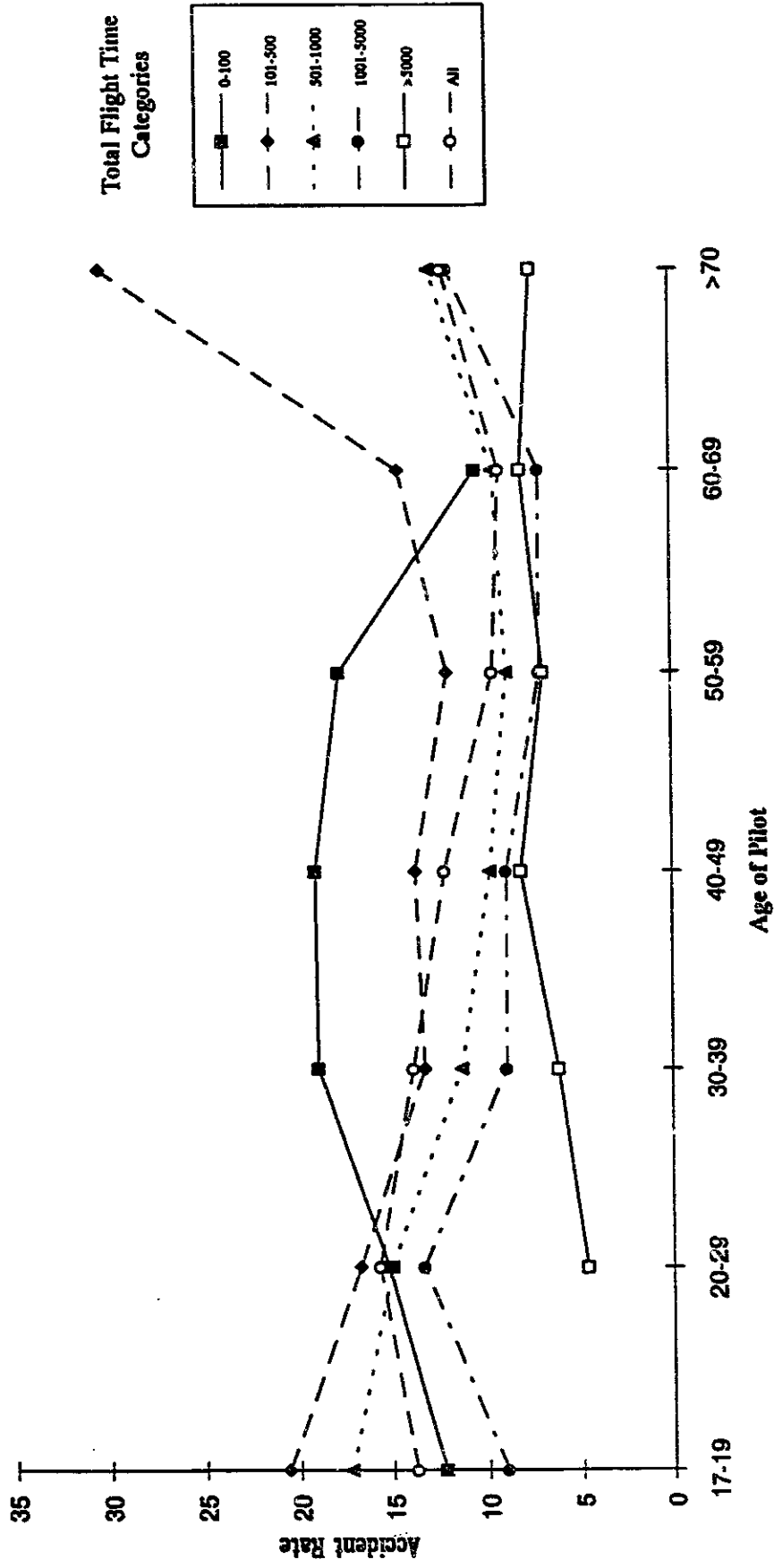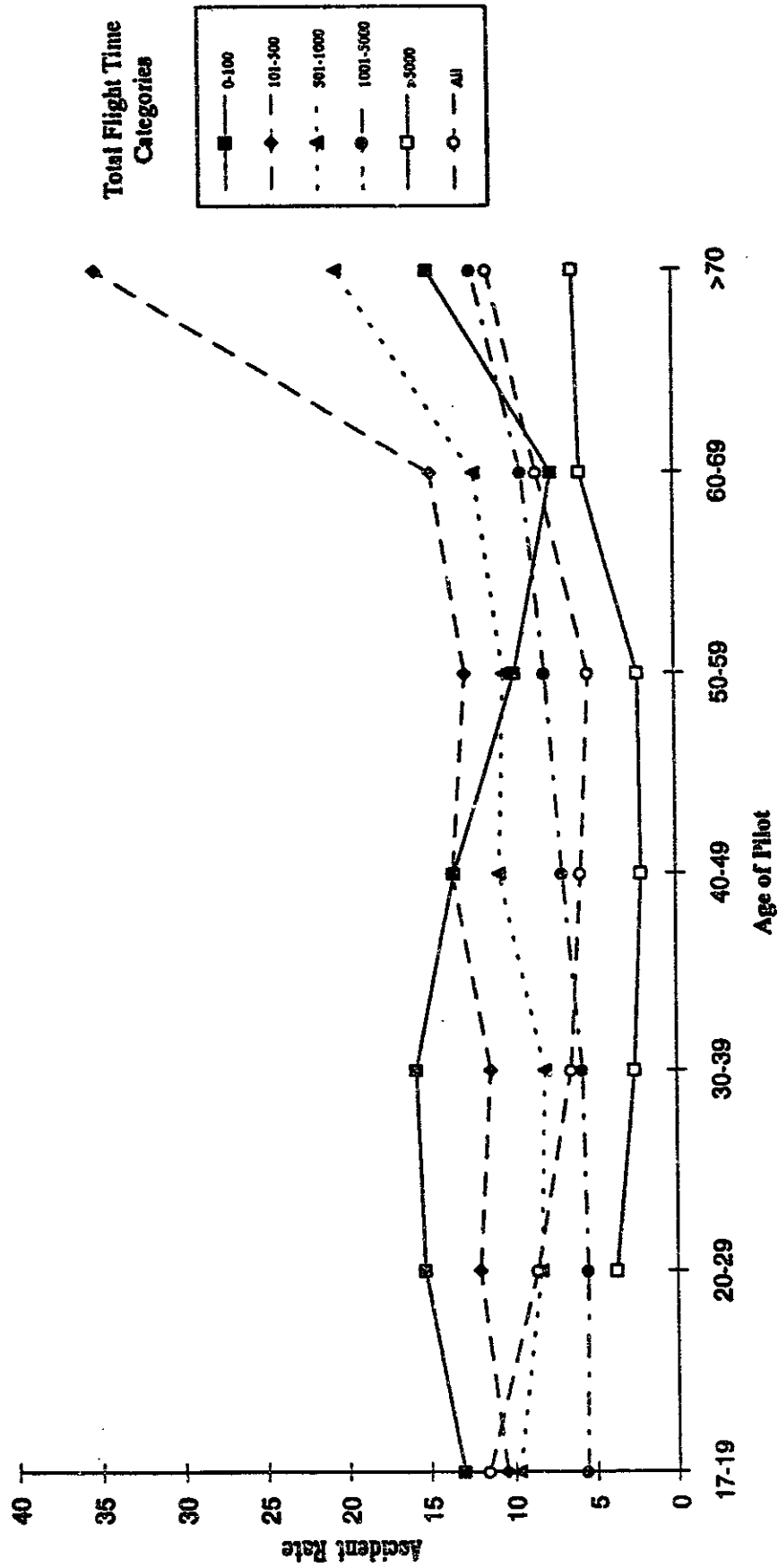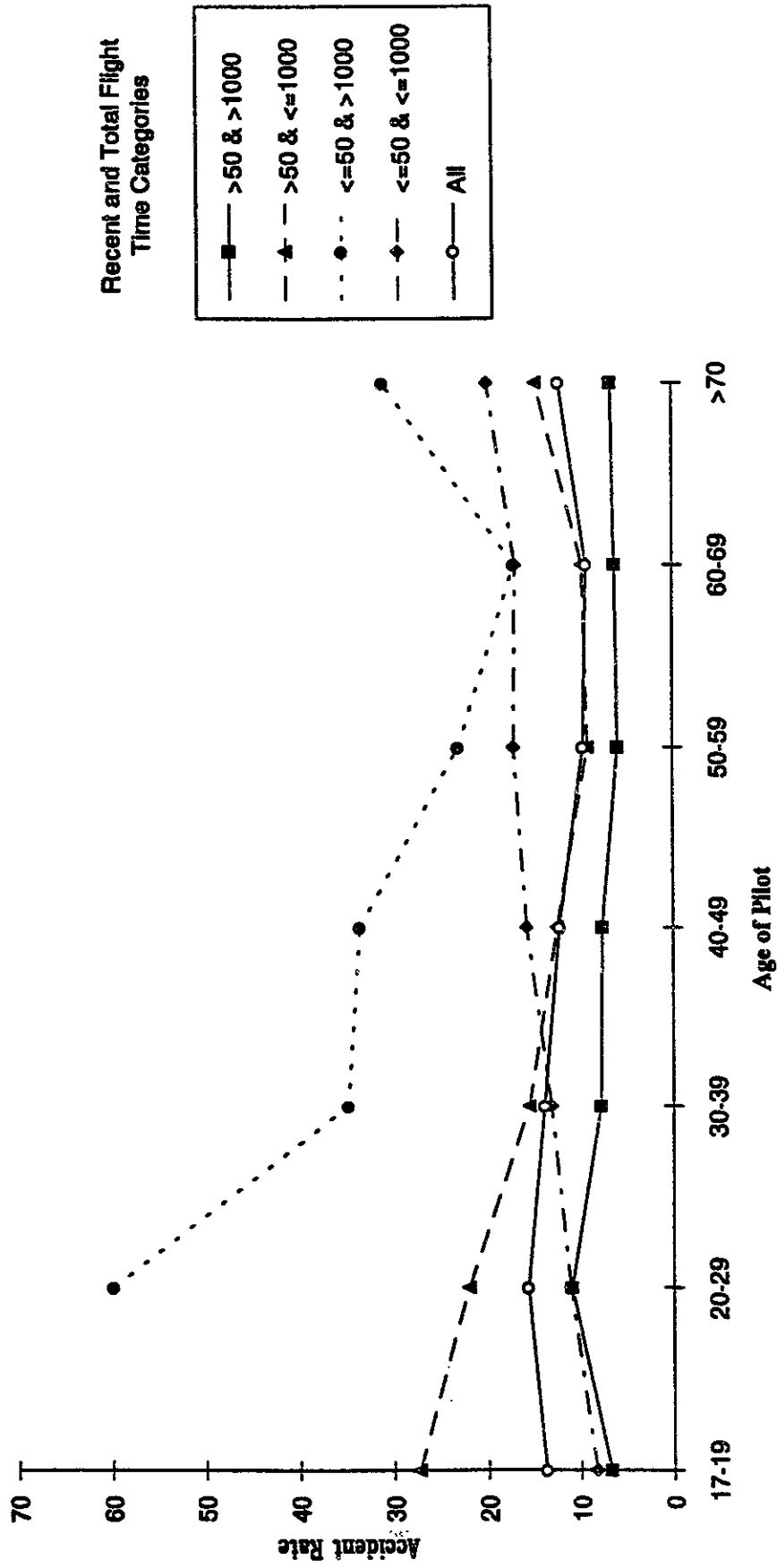1976-1980 Class I,II, and III Pilots
Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 4)

1976-1980 Class III Pilots
Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 5)

1976-1980 Class I, II, and III Pilots
Accidents per 100,000 Recent Flight Hours
Gołaszewski (Figure 6)

1976-1980 Class III Pilots
Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 7)

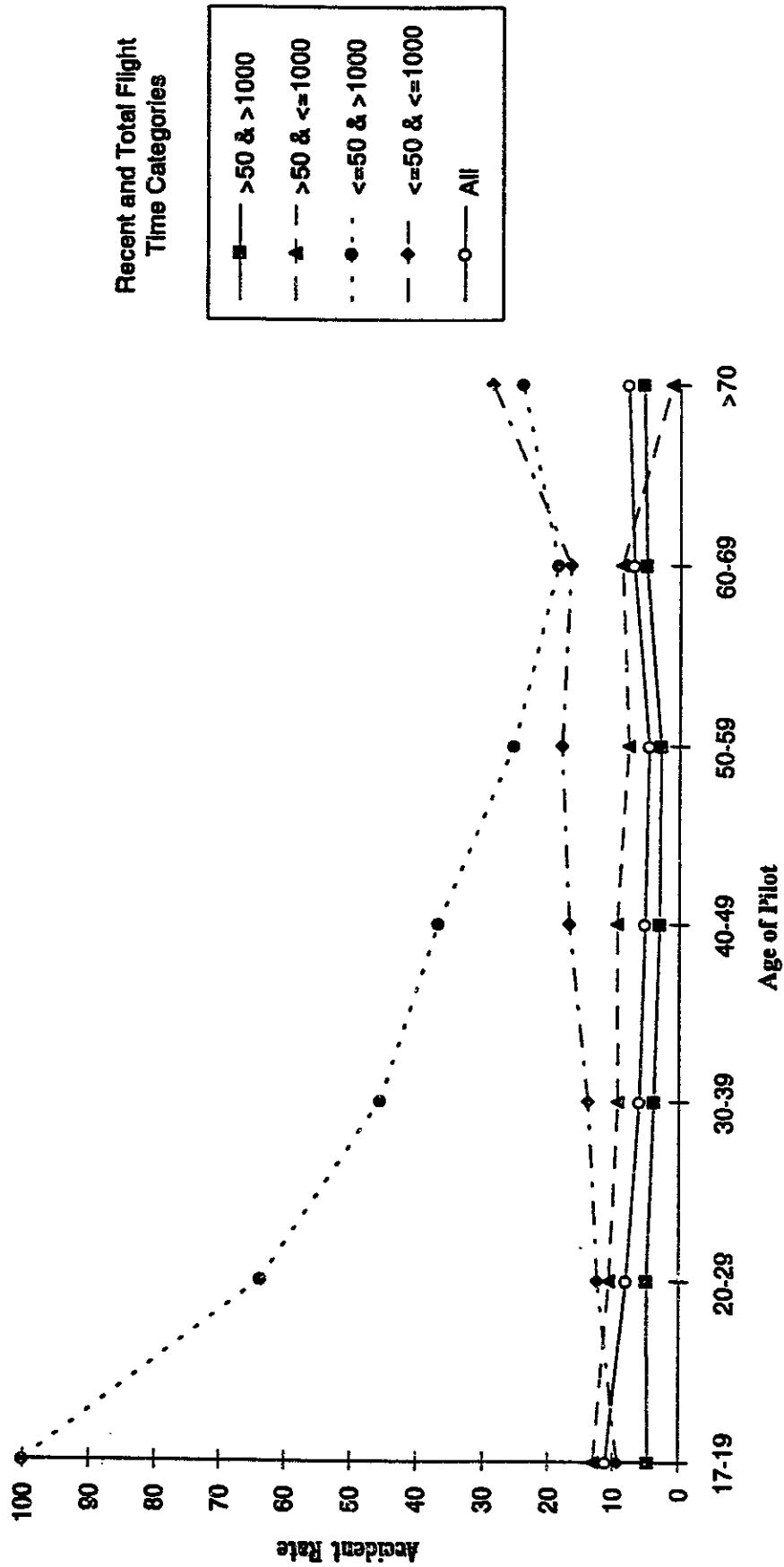1976-1980 Class I,II, and III Pilots
Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 8)

1976-1980 Class III Pilots
Accidents per 100,000 Recent Flight Hours
Golaszewski Figure 9

1976-1980 Class I, II, and III Pilots
Accidents per 100,000 Recent Flight Hours
Golaszewski (Figure 10)

APPENDIX C

Topical Bibliography of Aging, Medical Variables, and Pilot Performance

Edwin J. Kay
Lehigh University

CONTENTS
APPENDIX C

SECTION   **PAGE**

ACCELERATION TOLERATION

Suvorov, P. M. *Effect of age, occupation, and physical training on human tolerance to long-term acceleration.* Foreign Technology Div., Wright-Patterson AFB, OH. 16 Aug 89. 11p. Trans. of Komischeskaya Biologiya i Meditsina, (USSR) n2 p62-66 1968, by Marilyn Olaechea.

> The effect of the age, occupation and physical training on the human tolerance to long term accelerations was investigated. The lowest tolerance was found in the test subjects of the age range of 20-24 while the highest in those of 30-34. The test subject of 40-49 showed a decreased tolerance to accelerations.

ALCOHOL

Morrow, D., Leirer, V. & Yesavage, J. *The influence of alcohol and aging on radio communication during flight.* Aviat. Space Environ. Med. 1990 Jan. 61(1). P 12-20.

> Alcohol and age impaired communication-based and overall flying performance during and immediately after drinking. Alcohol and age cumulatively impaired performance, since older pilots were more impaired by alcohol. Moreover, overall performance was impaired for 8 h after reaching 0.10% BAC.

CARDIOVASCULAR EPIDEMIOLOGY

Booze, C. F., Pidkowicz, J. K., Davis, A. W., & Bolding, F. A. *Postmortem coronary atherosclerosis findings in general aviation accident pilot fatalities: 1975-77.* Aviat. Space Environ. Med. 1981 Jan. 52(1). P 24-7.

> Autopsies of 764 pilots involved in fatal general aviation accidents during the years 1975-77 were reviewed. Only 5% of the autopsied group had severe coronary atherosclerosis. The rate per 1,000 of severe coronary atherosclerosis increased with age from 14.5 for ages less than 30, to 89.9 for ages 50 years and above. The prevalence of coronary atherosclerosis among the autopsied airmen is less than for the general population.

Booze, C. F. Jr. & Staggs, C. M. *A comparison of postmortem coronary atherosclerosis findings in general aviation pilot fatalities.* Aviat. Space Environ. Med. 1987 Apr. 58(4). P 297-300.

> The autopsy reports of 710 pilots involved in fatal general aviation accidents for 1980-82 were reviewed to appraise the age-specific prevalence of coronary atherosclerosis. Only about 2.5% of the pilots had severe coronary atherosclerosis. Prevalence of severe coronary atherosclerosis increased with age from 5.8 per 1,000 for ages less than 40 years to 73.9 for age 50 years and above.

Bruce, R. A. & Fisher, L. D. *Exercise-enhanced risk factors for coronary heart disease vs. age as criteria for mandatory retirement of healthy pilots.* Aviat. Space Environ. Med. 1987 Aug. 58(8). P 792-8.

> Mortality due to coronary heart disease is examined in 4105 asymptomatic healthy men of 15 to 80 years of age observed in Seattle community practice. The proportion of men

of 60-64 years at low risk with a 98% 3-year survival rate is 16 times that of men of 55-59 years at high risk with a 78% survival rate.

Bugrov, S. A., Kiselev, R. K., Beleda, R. V., Plakhatnyuk, V. I. & Artamonov, N. N. *Blood Lipids and incidence of lipemia in flight personnel*. Space Biology and Aerospace Medicine, Vol. 20, No. 3, 1986 p 39-43.
> The content of cholesterol and triglycerides was measured in healthy pilots. Over 1500 subjects were examined and age-related norms of the parameter were established.

Gray, G. W. *Cardiopulmonary screening for high-performance flying: selection and retention issues*. Advisory Group for Aerospace Research and Development, Neuilly-sur-Seine (France). 1987. 4p.
> Detection methods used in the cardiopulmonary screening of aircrew are discussed with reference to candidate selection standards and retention of older experienced pilots.

Harrison, W. K. & Smith, J. E. *Age trends in the cardiovascular dynamics of aircrewmen*. Aviat. Space Environ. Med. 1979 Mar. 50(3). P 271-4.
> Age trends in the cardiovascular dynamics of 51 aircrewmen were investigated by means of noninvasive techniques. A statistically significant (p less than 0.01) decline in the ability of their hearts to accelerate blood was found to accompany advancing age.

Hoiberg, A. *Cardiovascular disease among U.S. Navy pilots*. Interim rept. Naval Health Research Center, San Diego, CA. Jul 84. 17p.
> Results showed a direct relationship between cardiovascular disease incidence and age.

Hoiberg, A. *Cardiovascular disease among U. S. Navy pilots*. Aviat. Space Environ. Med. 1985 May 56(5), 397-402.
> Aim to determine influence of age and occupation on cardiovascular disease (CVD) in Navy pilots diagnosed with CVD during 12.5 year period (N=150). CVD rate increased as a function of age. CVD onset occurred, on the average, 3 years earlier for pilots than for officers.

Hoiberg, A. *Longitudinal study of cardiovascular disease in U.S. Navy pilots*. Interim rept. Naval Health Research Center, San Diego, CA. Feb 85. 16p.
> This longitudinal study examined the consequences of cardiovascular disease (CVD) in 145 U.S. Navy pilots who suffered a CVD incident during 1967-79. The findings reflected not only the few CVD cases in this population of 22,245 pilots who served for some time from 1967-79, but also the few after-effects of CVD.

Hoiberg, A. & Blood, C. *Age-specific morbidity among Navy pilots*. Aviat. Space Environ. Med. 1983 Oct. 54(10). P 912-8.
> This study compares the morbidity (hospitalization) rates by age of male Navy aviators (n = 22,417) with rates for three male control populations: nonpilot aircrew officers (n = 9,483), unrestricted line officers (n = 55,593), and staff officers (n = 46,565). Aircrew members and pilots have the highest hospitalization rates. The highest rates for older pilots is for circulatory diseases.

Hoiberg, A. & Burr, R. G. *Health risks associated with aircraft model type among U.S. Navy pilots.* Proceedings of the Symposium: Psychology in the Department of Defense (9th). Colorado Springs, Colorado, 18-20 April 1984.

This study identified the ill health effects (hospitalizations) among U.S. Navy male pilots (n = 22,245) who primarily flew one of eight aircraft model types for July 1967 to December 1979. Older track pilots had the highest rates for accidental injuries, musculosketetal conditions, and infective/parasitic diseases. The highest circulatory disease rate was observed for older fighter pilots.

Jonsson, F. & Sundgren, N. *Essential hypertension in airline pilots.* Aerosp. Med. 1969 Jan. 40(1). P 70-5.

Affect of age considered.

Proper, R. *Preliminary impressions of cardiovascular function in the study of physiologic and psychologic aging in professional pilots.* Aerosp. Med. 1969 May. 40(5). P 557-60.

Proper, R. *Age-related changes in professional pilots as defined by the Klensch-Schwarzer ultra-low frequency ballistocardiogram.* Bibl. Cardiol. 1968. 20. P 50-6.

Scarborough, W. R., Westura, E. E. & Podolak, E. *Studies on a small group of normal test pilots with special reference to quantitative relationships of age with the ULF ballistocardiogram and other cardiovascular measurements.* Bibl. Cardiol. 1969. 24. P 66-74.

Underwood-Ground, K. E. *Prevalence of coronary atherosclerosis in healthy United Kingdom aviators.* Aviat. Space Environ. Med. 1981 Nov. 52(11 Pt 1). P 696-701.

The coronary arteries of 288 aircrew (135 military aircrew, 53 professional pilots, and 100 private pilots) killed in 210 aircraft accidents were examined to assess the prevalence of coronary artery disease (CAD) in four groups in a 2 factor (age and professional vs private pilot) design with mean ages 29.1, 29.9, 39.7, and 37.2. There was no difference in CAD prevalence in the four groups (p greater than 0.1).

York, E., Mitchell, R. E. & Graybiel, A. *Cardiovascular epidemiology, exercise, and health: 40-year followup of the U.S. Navy's "1000 aviators".* Aviat. Space Environ. Med. 1986 Jun. 57(6). P 597-9.

Beginning in 1940, a study of 1,056 student and instructor pilots lowered previously high attrition rates in training by emphasizing both physical and psychological screening. After World War II, when 208 pilots in the group died, followup studies of the survivors were conducted in 1951, 1957, 1963, 1969-71, 1977, and 1980-81. In February 1981, 715 questionnaires were mailed to known survivors, with 500 replies subsequently analyzed. Additionally, 114 of the respondents who had previously been examined during 1969, were again examined in 1980-81; those individuals were markedly different in their lifestyle, particularly in exercising regularly, abstaining from cigarette smoking, and drinking alcoholic beverages moderately, as contrasted to 28 aviators also examined in 1969 who died in the interim. Healthy lifestyle may alter cardiovascular risk, preventing premature death.

EPIDEMIOLOGY OF DISQUALIFICATION

Dark, S. J. *Characteristics of medically disqualified airline pilots.* Federal Aviation Administration, Washington, DC. Office of Aviation Medicine. Jan 83. 18p.
> Denial rates for airline pilots increase to the highest rate at age interval 55-59. The most significant causes for denial are cardiovascular, neuropsychiatric, and the miscellaneous category.

Dark, S. J. *Medically disqualified airline pilots.* Final rept. Performer: Federal Aviation Administration, Washington, DC. Office of Aviation Medicine. Jun 86. 18p.
> Rates of airline pilots denied medical certification in 1983 and 1984. Denial rates for airline pilots increase to the highest rate at age interval 55-59. The most significant causes for denial are cardiovascular, neuropsychiatric, and the miscellaneous category.

Whitton, R. C. *Medical disqualification in USAF pilots and navigators.* Aviat. Space. Environ. Med. 1984, 55(4), 332-336.
> Displays tables of medical condition and age of 304 USAF personnel disqualified during 1980 and 1981. Cardiovascular disease was most common cause (30%) for disqualification.

FITNESS

Nunneley, S. A., Finkelstein, S. & Luft, U. C. *Longitudinal study on physical performance of ten pilots over a ten-year period.* Aerosp. Med. 1972 May. 43(5). P 541-4.

GENERAL CARDIOVASCULAR

Durrer, D. *Cardiological aspects of the aging pilot.* Aeorspace Med. 1974, 45(4): 438-442.
> Summary of panel discussion of Airlines Medical Directors Association meeting, May 5, 1973, in Las Vegas.

Loeppky, J. A. & Luft, U. C. *Work capacity, exercise responses and body composition of professional pilots in relation to age.* Aviat. Space Environ. Med. 1989 Nov. 60(11). P 1077-84.
> Body composition and cardiorespiratory responses during a progressive upright bicycle ergometer test were measured in 410 professional male pilots, aged 20 to 68 years. Aerobic work capacity (VO2max) fell at a rate of 0.25 ml.min-1.kg-1 per year in this unique population of healthy, but generally sedentary men.

Mertens, H. W., Higgins, E. A. & McKenzie, J. M. *Age, altitude, and workload effects on complex performance.* Federal Aviation Administration, Washington, DC. Office of Aviation Medicine. Sep 83. 18p.
> Fifteen healthy men in each of three age groups, 20-29 yrs, 40-49 yrs, and 60-69 yrs, were evaluated regarding complex performance in two altitude conditions (ground level vs. 3,810 m). Heart rate decreased slightly at the 3,810 m altitude in the 60-69 yr group, but increased significantly at altitude in the two younger groups. Both

epinephrine and norepinephrine excretion rates were highest in the 20-29 yr group and lowest in the 40-49 yr group.

Scheinman, H. Z. *Coronary atherosclerosis in military pilots. I.* Relationship to flying and aviation accidents. Aerosp. Med. 1968 Dec. 39(12). P 1348-51.

Szafran, J. *Age differences in the rate of gain of information, signal detection strategy and cardiovascular status among pilots.* Gerontologia. 1966. 12(1). P 6-17.

GENERAL EPIDEMIOLOGY

Burr, R. G. & Hoiberg, A. *Health profile of U.S. Navy pilots of electronically modified aircraft.* Aviat. Space Environ. Med. 1988, 59:168-171.
> Compared hospitalization rates of pilots of electronically modified aircraft (n=1063) with age-matched sample of pilots of other aircraft (n=2126). Increases of cardiovascular disease and alcoholism as a function of age for control group but not experimental group.

Guide, P. C. & Gibson, R. S. *An analytical study of the effects of age and experience on flight safety.* Proc. Human Factors Soc. 35th annual meeting, 1991, 180-184.
> Authors plot accidents per number of pilots and accidents per hours flown as a function of age. Roughly a replication of Golaszewski, with the same limitations.

Holt, G. W., Taylor, W. F. & Carter, E. T. *Airline pilot medical disability: A comparison between three airlines with different approaches to medical monitoring.* Aviat. Space Environ. Med. 1987, 58:788-791.
> Cohort of 12,866 pilots studied from 1975 to 1982. Medical losses increased rapidly after pilots reach 45, with cardiovascular disease responsible for about half the losses.

Vlasov, V, V. & Kopanev, V. I. *Effect of flying on the health status of pilots in the light of epidemiologic data.* Kosm. Biol. Aviakosm. Med. 1990 Jan-Feb. 24(1). P 4-9. (Review).
> This paper gives an analytical review of epidemiological data about the morbidity and mortality rate of the flying personnel. Flying personnel differ from the general population in their better health condition which manifests as lower morbidity and mortality rates in every large group of disease. This better health maybe a result of their selection, which makes it difficult to identify relationships between disease and flight effects.

HEARING

Fitzpatrick, D. T. *An analysis of noise-induced hearing loss in Army helicopter pilots.* Aviat. Space Environ. Med. 1988 Oct. 59(10). P 937-41.
> Hearing loss is primarily a function of noise exposure as measured by total flight hours. Age was found to be a less significant factor; aircraft type had no significant effect.

Schulthess, G. von. *Statistical evaluation of hearing losses in military pilots.* Acta Otolaryngol. (Stockh). 1969 Sep. 68(3). P 250-6.

NEUROLOGICAL

Marks, E., Zuzhevich, V., Dvoretski, E. & Mazhentski, M. *Age-related changes in the electroencephalograms of pilots.* Kosm. Biol. Aviakosm. Med. 1985 Sep-Oct. 19(5). P 85-6.

PREDICTION OF INCAPACITATION

Boone, J.O. *Functional aging in pilots: an examination of a mathematical model based on medical data on general aviation pilots.* Federal Aviation Administration, Washington, DC. Office of Aviation Medicine. 1982.
> Age is not as accurate in discriminating between the sudden incapacitating pathology and nonpathology groups.

Booze, C. F., Jr. *Sudden inflight incabacitation in general aviation.* Aviat. Space Environ. Med. 1989, 60, 332-335.
> Study of CAMI data for 1975-1982. About 3 accidents per 1000 due to incapacitation. Results suggest the likelihood of incapacitation increases with age.

Castelo-Branco, A., Cabral, Sa. A. & Coelho-Borges, J. *Comparative study of physical and mental incapacities among Portuguese airline pilots under and over age 60.* Aviat. Space Environ. Med. 1985 Aug. 56(8). P 752-7.
> The number of definitive flight incapacitations and deaths among Portuguese airline pilots between 1945 and 1983. The rate of incapacities became higher than under age 60, but 64% of over 60 examinees were absolutely fit for flight duties.

Froom, P., Benbassat, J., Gross, M., Ribak, J. & Lewis, B. S. *Air accidents, pilot experience, and disease-related inflight sudden incapacitation.* Aviat. Space Environ. Med. 1988 Mar. 59(3). P 278-81.
> It may be preferable to grant waivers to experienced pilots with an increased incidence of disease-related inflight sudden incapacitation than to replace them with novices. Overly strict medical criteria may paradoxically increase accident rates.

Kulak, L. L., Wick, R. L. & Billings, C. E. *Epidemiological study of airline pilot incapacitation.* 1971, Aeorspace Med. 42(6):670-672.
> Rates of career termination during 11 year period determined as a function of age. Rates increase with age, with coronary heart disease playing a major role. Age-specific incidence of heart disease lower in pilots than in general population.

Lane, J. C. *Risk of in-flight incapacitation of airline pilots.* Aerosp. Med. 1971 Dec. 42(12). P 1319-21.

Martin-Saint-Laurent, A., Lavernhe, J., Casano, G. & Simkoff, A. *Clinical aspects of inflight incapacitations in commercial aviation.* Aviat. Space ENviron. Med. 1990, 61, 256-260.

Studied causes of inflight incapacitation in Air France pilots from 1968-1988. None caused an accident. Incapacitations of a cardiac nature were rarer and less serious than those caused by gastrointestinal or neurological disorders.

Rayman, R. *Myocardial infarction: an in-flight problem?* Aerospace Med: 1974, 45(1):86-89.
    In-light myocardial infarction has been a rare event during 1962-1971 in the USAF.

Rayman, R. & McNaughton, G. *Sudden incapacitation: USAF experience, 1970-1980.* Aviat. Space Environ. Med., 1983, 54(2):161-164.
    Authors examine 146 reported cases of sudden in-flight incapacitation in the USAF. Of the 17 fatalities, 1 was suspected to be from myocardial infarction.

Tunstall-Pedoe, H. *Risk of a coronary heart attack in the normal population and how it might be modified in flyers.* Eur. Heart J. 1984 Mar. 5 Suppl A. P 43-9. (Review).
    About 50% of heart attacks might cause immediate total incapacitation, rates are such that only a very small percentage of fatal accidents could come from this cause.

RESPIRATORY DISEASES

Catlett, G. F. & Kidera, G. J. *Detection of respiratory impairment in pilots.* Aeros. Med. 1969 Nov. 40(11). P 1252-7.

Lategola, M. T., Flux, M. & Lyne, P. J. *Spirometric assessment of potential respiratory impairment in general aviation airmen.* Aviat. Space Environ. Med. 1977, 48(6):508-511.
    Spirometric measurements of 181 male general aviation pilots showed increasing spirometric impairment with age and with amount of cigarettes smoked.

Voge, V. M. & Anthracite, R. *Spontaneous pneumothorax in the USAF aircrew population: a retrospective study.* Aviat. Space Environ. Med. 1986, 57:939-949.
    Survey of all aircrew in USAF waiver file. The absolute number of occurrences of spontaneous pneumothorax decreases with age.

VESTIBULAR

Moser, M. H. & Ranacher, G. R. *Medical fitness examination of commercial pilots: new criteria for evaluation of vestibular test.* Aviat. Space Environ. Med. 1982, 53(12) 1215-1219.
    Compensation capacity of the eyes (to rotation of the head) decreases as a function of age.

VISION

Baisden, A. G., Monaco, W. A. (1983). *Distribution of Visual Characteristics of Naval Aviation Personnel.* Final Report. Naval Aeorospace Medical Research Lab, Pensacola, FL.
    With increasing age there are decreases in visual acuity accomodative amplitude, increases against-the-rule astigmatism and myopia, and stability in fusion-related variables and intraocular pressure

Elworth, C. L., Larry, C. & Malmstrom, F. V. *Age, degraded viewing environments, and the speed of accommodation.* Aviat. Space Environ. Med. 1986 Jan. 57(1). P 54-8.

> The time to accommodate from a near to infinity target varied as a function of age, room illumination, and the distance of the near-reading task. During ideal conditions, older subjects could accommodate nearly as quickly as the younger subjects; however, during degraded viewing conditions, the accommodation time for older subjects increased as much as tenfold.

Morris, A. & Temme, L. A. *Time Required for U.S. Navy Fighter Pilots to Shift Gaze and Identify Near and Far Targets.* Aviation, Space, and Environmental Medicine, 1989, v60 p1085-1089.

> The speed with which 163 U.S. Navy fighter pilots can shift their line of sight and discriminate high contrast targets was measured. Both Far-to-Near and Near-to-Far times significantly slowed with age, the difference due to the oldest subject (44 years of age).

Sekuler, R., Kline, D. & Dismukes, K. *Aging and Visual Function of Military Pilots: A Review.* National Research Council, Washington, DC. Committee on Vision.

Temme, L. A. & Morris, A. *Speed of Accommodation and Age.* Optometry and Vision Science, v66 n2 p106-112 Feb 89.

> The time needed to change accommodation was measured in 65 U.S Navy fighter pilots whose age ranged from 24 to 44 years. The speed of accommodative change, far-to-near (FN), slowed with age but near-to-far (NF) did not.

Vola, J. L., Cornu, L., Carruel, C., Gastaud, P. & Leid, J. *Age and photopic and mesopic visual acuity.* J. Fr. Ophtalmol. 1983. 6(5). P 473-9.

> Visual acuity at high luminance (100 cdm2) and at low luminance (0.8 cdm2) was measured in normal subjects aged between 20 and 50 years. Low luminance acuity decreased with age whereas high luminance acuity remained unchanged (20/20).