DOT/FAA/AM-94/23

Office of Aviation Medicine
Washington, D.C. 20591

Age 60 Study, Part IV: Experimental Evaluation of Pilot Performance

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

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

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

October 1994

Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

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.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-94/23		

4. Title and Subtitle	5. Report Date
Age 60 Rule Research, Part IV: Experimental Evaluation of Pilot Performance	October 1994
	6. Performing Organization Code
7. Author(s) D. T. Hyland, Ph.D., E. J. Kay, Ph.D., J. D. Deimler	8. Performing Organization Report No. 8025-4B
9. Performing Organization Name and Address Hilton Systems, Inc. 950 North Kings Highway Suite 202 Cherry Hill, New Jersey 08034	10. Work Unit No. (TRAIS) 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. Supplemental 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 under Task AM-A-91-HRR-138.

16. Abstract

This report was a deliverable from the research contract with Hilton Systems, Inc. on the FAA's mandatory retirement for pilots operating under Federal Aviation Regulations Part 121, the "Age 60 Rule." The purpose of this study was to examine the feasibility of developing an individually-based pilot performance assessment, as well as design an experimental methodology to empirically examine the relationship between pilot aging and performance. Pilot performance was measured with both domain-dependent, as well as domain-independent assessments to test a decrement with compensation model of expertise and aging. Computerized cognitive test batteries, COGSREEN and WOMBAT, were selected as the domain-independent measures. Flitescript and whole task performance in the B727 simulator were domain-dependent measures. Forty B727-rated pilots were recruited from air carriers and the FAA. Pilots were males between the ages of 41 and 71 years (M=53.9, sd=8.1). All pilots had a minimum of 5000 hours of total flight time with a wide range of total and recent hours in type. Three simulator scenarios were designed to assess pilot performance on routine and emergency/abnormal maneuvers. Simulator performance measures were based on a deviation score and an evaluator rating. The relationships between the following measures were assessed by examination of the correlations between: 1) flying experience and simulator performance, 2) predictor test scores and simulator performance, 3) interrelationships between the predictor tests, and 4) age, flying experience, predictor test scores and simulator performance. Finally, pilot perceptions of each measure were assessed. COGSCREEN total composite scores were significantly correlated with evaluator ratings on emergency/abnormal maneuvers. Neither WOMBAT nor Flitescript were found to correlate with simulator performance. Pilot age was significantly correlated with performance on the predictor tests. A pattern of inter-correlations among pilot age, COGSCREEN and simulator performance was discussed.

17. Key Words	18. Distribution Statement
Pilot age and performance Age 60 Rule Pilot performance	Document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	65	

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

PREFACE

This technical report, entitled "Age 60 Rule Research, Part IV: Experimental Evaluation of Pilot Performance," is the fourth document in the series of products published as technical reports from the Age 60 Rule research contract with Hilton Systems 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 second report was published as "Age 60 Rule Research, Part II: Airline Pilot Age and Performance--A Review of the Scientific Literature" (DOT/FAA/AM-94/21). The third report was entitled "Age 60 Rule Research, Part III: Consolidated Database Experiments Final Report."

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 1960 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.

The contract required a second scientific approach to mandatory pilot retirement to supplement the data-based analyses of accident data (reported in "Age 60 Consolidated Database Experiments Final Report"). This contract requirement resulted from recommendations in a 1981 Institute of Medicine report which recommended a systematic program to collect medical and performance data, necessary to support adequate assessment of the Age 60 Rule, be initiated. An associated report from the National Institute on Aging (1981) recommended that additional research be conducted to develop quantifiable, objective criteria for measuring overall pilot performance. This report presents the initial results from the contract research.

This project was a collaborative effort between the contractor, Hilton Systems, their consultants, Lehigh University, and the FAA. I would personally like to acknowledge Mr. A.E. Dillard, FAA National Resource Specialist, Simulator Engineering, for his dedicated participation on this project involving assistance with simulator time, scenario development, recruiting subjects and staff, and coordinating contacts with industry. In addition, Mr. Al

Hendrix contributed many hours toward developing simulator psychometric measures and scenario protocols, as well as onsite project management. I am grateful for their assistance.

Pamela Della Rocco, COTR

Hilton Systems, Inc. and Lehigh University Personnel

Mr. Albert Zalcman, Program Manager
Mr. Jim Deimler, Principal Engineer
Dr. Donald Hillman, Chairman of the Computer Science and Electrical Engineering Department (CSEE)
Dr. Edwin Kay, Faculty member of CSEE Department
Dr. Diane Hyland, Director of the Center for Social Research, Lehigh University
Mr. Nelson Brown, ret. Eastern Airlines pilot
Ms. Julie Chambliss, Hilton Systems
Ms. Regina Harris, Hilton Systems
Mr. Robert Voros, Lehigh University

Members of Scientific Group of Experts

Susan Baker, M.P.H., The Johns Hopkins University
Donald Campbell, Ph.D., Lehigh University
James Danaher, National Transportation Safety Board
Russell Raymond, M.D., Aerospace Medical Association
Ilene Siegler, Ph.D., Duke University;

FAA Participants

Garnett "Mac" MacLean, Ph.D.
Carol Manning, Ph.D.
Steve Veronneau, M.D.
Henry Mertens, Ph.D.
Robert Blanchard, Ph.D.
Mr. A.E. Dillard, National Resource Specialist, Simulator Engineering
Mr. Anthony Broderick (AVR-1)
Dr. Jon Jordan, Federal Air Surgeon (AAM-1)

ACKNOWLEDGMENTS

This research was supported by the FAA Civil Aeromedical Institute (CAMI) in Oklahoma City, Oklahoma, under contract No. DTFA-02-90-90125.

The authors acknowledge a number of individuals for their assistance during this research. We thank Archie Dillard for the use of the B727 simulator and for invaluable assistance throughout the planning and implementation of the simulator evaluation. We thank Nelson Brown for assistance in scenario development and evaluator training. We thank Al Hendrix for assistance in scenario development and for the many hours spent as test administrator. We thank Helene Iavecchia for assistance in the development of the test plan. We thank Crystal Cruz and Otto MacLin of CAMI for assistance with the administration of the test battery.

We thank the pilot evaluators, George Budde, Bill Duvall, Bill Klark, and Ed Stevens, as well as Jim Nebgen for his skillful operation of the simulator. We especially thank the 40 pilots who volunteered their time as test subjects and who will remain anonymous. Their willingness to support this important research is appreciated.

Finally, and foremost, the authors thank the CAMI contract monitor, Pam Della Rocco, for invaluable guidance, assistance, and understanding throughout the duration of the research.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

1.0 INTRODUCTION

1.1 Background

The "Age 60 Rule", contained in Part 121 of the Federal Aviation Regulations (121.383C), mandates retirement at age 60 for 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. 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. While there was recognition of the fact that not all individuals experience equivalent age-related deteriorations in health and performance, it was concluded that an age limitation was 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) 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 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, they made two major recommendations:

- 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.

In 1990, the Office of Technology Assessment (OTA) conducted another evaluation of the Age 60 Rule. This report also concluded that the scientific evidence was inadequate to resolve the Age 60 question, and therefore the report suggested that further research on pilot age and performance be conducted before considering a change to the Age 60 Rule.

In 1990, a team of researchers from Lehigh University and Hilton Systems began working under a two year FAA contract to Hilton Systems, Inc. on a series of tasks concerning aging, pilot performance and the Age 60 Rule. The contract was managed through the Civil Aeromedical Institute (CAMI). The major thrust of the Age 60 Rule contract was the consolidation of several FAA and NTSB databases and a series of analyses that explored the relationships among pilot age, pilot flying experience and accident rates (Kay, Harris, Voros, Hillman, Hyland, & Deimler, 1992). The Age 60 Rule research contract study also included an exhaustive review of the scientific literature in the area of aging and pilot performance and development of a conceptual model to guide research on aging and pilot performance (Hyland, Kay, Deimler, & Gurman, 1992). This report describes the third major task of the Age 60 Rule Study: the development of a methodology to quantitatively, objectively and comprehensively assess an individual pilot's performance and the preliminary examination of the relationship between aging and pilot performance.

1.2 Objectives

The long-term aim of this research is to increase understanding about the relationships among pilot age, experience, and performance. This information is critical to making informed decisions about the Age 60 Rule. As the next section will describe, there is surprisingly little data available about developmental changes in pilot performance. There is a critical need to collect comprehensive, objective performance data from an appropriately selected sample of pilots who vary in chronological age.

Further examination of the influence of increasing age on pilot performance must procede in two interrelated directions. The first direction focuses on interindividual differences, the variability among individual pilots' performance or abilities. Here the critical question arises of whether it is possible to objectively, reliably, and validly assess a particular pilot's performance with sufficient confidence that the correct employment decisions are made. This is the type of information which is critical to determining the feasibility of an individual performance-based standard as a replacement to the current chronological age standard. Both the MA and OTA reports concluded that there were no existing measures, that could serve as individualized assessment approaches, that have been validated against reliable and relevant measures of pilot performance. They were also aware of no available standards for grading pilot performance in a simulator in objective, quantifiable ways. Thus, a major goal of the present project is the development of pilot assessment procedures.

Thorough re-evaluation of the Age 60 Rule must also move in a second direction which involves a focus on age differences and changes. Here the focus shifts from assessing the performance of particular pilots to a search for normative developmental patterns for a group of pilots. Now the critical questions become: What is the typical pattern of change in pilot performance across a particular age span? How much variability is there in this pattern of change? At what age is change in performance most likely to occur? The information obtained from this developmental approach is critical for evaluating age-based group standards. Is age 60 the best age for this aviation regulation or does data more strongly indicate another age? These two approaches are interrelated because a reliable, valid method of assessing individual pilot performance is required to examine developmental age differences and changes in pilot performance.

Thus, the specific research objectives of the present project were:

1. Development of a methodology to quantitatively, objectively and comprehensively assess an individual pilot's performance. Specifically,

measures at each of these three levels of performance were investigated:

- Domain-independent psychomotor, perceptual and cognitive skills which were not specifically related to flying but were selected because of their relevance to performance;

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

- Complex whole task performance in a simulator under varying flight conditions.

2. Examination of relationships among these performance measures in a group of pilots varying in age. The data obtained were used to:

   - Examine the relationship between the predictor variables of domain-independent skills and domain-dependent knowledge and the criterion of performance in the simulator.

   - Examine the relationships among the predictors to reduce redundancy and provide the most concise battery of predictor tests.

   - Examine the ease of use and pilot acceptance of these measures as an individual performance assessment standard that could eventually replace the Age 60 Rule.

3. Preliminary examination of relationships between pilot age and these three levels of pilot performance. Key developmental questions addressed were:

   - With increasing age were there declines in performance on measures of basic skills, pilot knowledge and decision making, or simulated flying performance?

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

To accomplish the primary research objectives of this project the following steps were performed:

1. Reviewed and analyzed gerontology and pilot performance literature to identify key performance variables and existing measurement methods.

2. Developed a high-level model for assessing aging, experience, and pilot performance.

3. Developed an objective quantifiable criterion measure of complex pilot performance in the simulator.

4. Constructed a test battery, consisting of available measurement techniques, that assessed pilot cognitive abilities, skills and knowledge.

5. Collected data on pilots which included the test battery and the simulator criterion measures.

6. Analyzed and reported the results.

The results of steps 1 and 2 are reported in Hyland, Kay, Deimler and Gurman (1992) and summarized in Section 2 of this report. The results of steps 3, 4, and 5 are described in Section 3. Step 6 is reported in Section 4.

2.0 CONCEPTUAL APPROACH

2.1 Review of the Literature on Aging and Pilot Performance

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. Hyland, Kay, Deimler, and Gurman (1991) reviewed the aviation human factors literature to identify the cognitive functions considered to be crucial to piloting. Existing conceptualizations of the abilities underlying pilot performance (Braune & Wickens, 1984a; Gerathewohl, 1977; 1978a; 1978b; Imhoff & Levine, 1981; North & Gopher, 1976) emphasized processing speed, attention, psychomotor skills, perceptual abilities and memory. They then reviewed the cognitive aging literature to determine the degree to which each of these functions is affected by aging and/or experience. Unfortunately, firm conclusions about most of these constructs were impossible to draw because of the shortcomings in the literature described below.

With regard to the literature on human factors aspects of piloting aircraft, there is a large and burgeoning literature, yet comparatively little aimed directly at the effect of aging on pilot performance. 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. The second category of studies (Braune & Wickens, 1984a; 1984b; Szafran, 1966; 1969) involved assessment of a range of psychological abilities presumed to be important to pilot performance and correlation of such data with chronological age and pilot performance. For example, 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 psychophysiological skills. Each pilot was administered a battery of tests that was designed to measure the following flight-related skills: making high-speed decisions, detecting low probability and low intensity signals, and retaining significant amounts of information (Szafran, 1966). The conclusion for almost all measures studied was that pilots' increasing age was not related to performance of these flight-related tasks.

The 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 inflight performance.

Because of the relatively small number of studies that have directly investigated the effects of aging on specific aspects of pilot performance, it is necessary to rely on studies within the area of cognitive aging that have examined non-pilot subjects performing more generic laboratory tasks. The available data on laboratory-based performance tests of basic perceptual, psychomotor, and cognitive skills suggest 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. Extreme care must 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 citizen 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). Most importantly, these laboratory-based performance tests have not yet been demonstrated to be predictors of piloting performance. 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.

Because of the relatively small number of studies that have directly examined aging pilots and the shortcomings described above it is impossible to determine from the existing literature the degree to which pilot performance is affected by aging and/or experience. Therefore, 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.

2.2 Conceptual Framework

The framework for the present project attempts to draw together the fragmented research on aging and pilot performance from the cognitive aging and aviation human factors literature into a model of aging and pilot performance. This model 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.

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. However, some researchers (Charness, 1983; Rybash, Hoyer, & Roodin, 1987; Salthouse, 1987) studied age differences in complex behaviors such as typing, chess, bridge, physics and computer programming and 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. For example, in the current context it is possible that domain-relevant knowledge and judgment related to flying may compensate for any observed declines in an older pilot's processing speed and working memory.

It is important to determine in what domains and to what degree expertise in terms of an enriched knowledge base 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 domains, such as chess or computer programming, where performance allows more time for predictability, planning ahead, and reflection, and demands fewer snap decisions and 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 most 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 may become more similar to "basketball" and compensation may no longer be sufficient to maintain the same level of complex task performance.

Thus, a decrement with compensation model applied to the study of aging and pilot performance suggests that data collection should be aimed at answering the following types of questions:

- With increasing age are there declines in performance on measures of 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? 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?

3.0 METHODOLOGICAL APPROACH

The primary research objectives of this project were to:

- Develop a criterion measure of complex pilot performance in a simulator that was objective and quantifiable.

- Develop a test battery of component skills and abilities that were assumed to contribute to complex performance. The test battery measurements served as predictor variables of simulator performance.

- Collect data on the predictor variables and the simulator criterion measure within a group of 40 pilots varying in age.

3.1 Development of the Simulator Performance Measure

The aviation community accepts simulator performance as the closest measure of a pilot's capability outside of the cockpit. High fidelity simulators have assumed a commanding role in the process by which Part 121 air carriers train pilots, certify their competence, and conduct periodic proficiency checks. 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. Currently, proficiency check methods used by Part 121 air carriers are qualitative in nature. The individual being evaluated flies a scenario that presents routine and non-routine situations.

Performance is observed by a 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 this qualitative assessment approach is subjective in nature and is subject to possible instructor bias, it is widely used in the industry and has proven itself to be operationally sound. However, more quantifiable measures of performance are needed in the research context. The quantitative techniques currently available are less developed and frequently oriented to military-related aviation, where performance requirements are imposed by the unique needs of combat (Kruk, Regan, Beverly, & Longridge, 1983). Although scientifically interesting, development in the military context is of limited use in studies of aging and commercial pilot performance.

The Lehigh/Hilton Systems research team worked with FAA personnel (including simulator experts, instructors and check pilots) to develop a simulator assessment procedure that was more objective, quantifiable, and definitive than the typical proficiency "check-ride." To develop the simulator criterion measure, the research design team completed the following steps:

- Developed three scenarios to be executed in the simulator that were realistic, challenging, and had high face validity among pilots.
- Selected relevant and credible maneuvers for inclusion in the scenarios which required both high and low workload.
- Defined the component actions required in each maneuver.
- Defined desired levels of performance for component actions within each maneuver (e.g., flight-parameter tolerance limits or pilot procedures).
- Identified specific mission points when flight parameters and or pilot actions will be rated by evaluators (and sometimes recorded by the simulator computer).
- Developed a composite index of performance for each maneuver.
- Developed a composite index of global performance.

Using this procedure, scoring protocols for three flight scenarios were developed (see Appendix A). Scenarios were constructed so that pilot proficiency could 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). Mertens, Higgins, and McKenzie (1983) found a significant interaction between age and workload. 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 a moderate or high workload level. Standard maneuvers, because they are so well-learned and practiced, may only be useful in detecting obvious decrements in pilot performance. Therefore the scenarios constructed for the present project incorporated routine maneuvers, abnormal situations, and emergency flight conditions.

Scenario 1 involved routine procedures carried out under low to moderate workload conditions. Pilots performed the following maneuvers: preflight checklist, engine start, taxi, takeoff, climbs, turns, descending turns, steep turns, level flight, tracking, precision approach and landing.

Scenario 2 involved routine and emergency procedures carried out under moderate to high workload conditions. Pilots performed the following maneuvers: checklist, engine start, IFR departure, aborted takeoff, normal takeoff, holding pattern, engine loss, non-precision approach, missed approach and VFR landing.

Scenario 3 involved routine and emergency procedures that emphasized pilot decision making carried out under moderate to high workload conditions. Pilots performed the following maneuvers: checklist, engine start, IFR departure, normal takeoff, weather which required new IFR clearance, holding pattern, engine loss on holding pattern entry, back course ILS approach, missed approach at MDA due to ground fog, second engine loss during missed approach, decision to land on any runway, and a single engine VFR landing.

Performance on the simulator scenarios was scored by evaluators using the following procedures. The pilot rater used a detailed maneuver scoring sheet which listed the objective criteria for each key action in that maneuver and records deviations from error free performance. These deviations could range from 1 (for a minor deviation) to 20 (for a major excursion from the appropriate action). For each subject, total deviations for each maneuver in each of the three scenarios were calculated. The pilot rater also assigned a subjective overall rating for each maneuver and each scenario using a 0 to 100 scale with 50 labeled as below average performance, 75 labeled as average performance and 100 labeled as error free performance. The pilot rater who served as a co-pilot for a subject also assigned a subjective overall rating for each scenario using the same 0 to 100 scale. A team of four check pilots were trained to serve as observer/raters. Training continued until observers met an acceptable criterion of interrater reliability for the deviation ratings and subjective evaluation ratings.

3.2 Construction of the Predictor Battery

The second task of the project was to construct a predictive test battery, based on individualized assessment of psychological functioning, which could be validated against the simulator criterion measure. Both the MA and OTA reports had concluded that there were no existing individualized assessment measures that had been validated against reliable and relevant measures of pilot performance. A number of aviation-relevant cognitive skills batteries have been constructed. However, most have been designed and validated for very specific military purposes, such as pilot selection (Carretta, 1987; Damos & Gibb, 1986). These batteries have typically been validated against a criterion like successful completion of flight training, not actual pilot performance in skilled pilots. There are other promising cognitive skills test batteries which have been developed to examine effects of aging or neurological impairment in pilots (Horst & Kay, 1991; Stokes, Banich, & Elledge, 1988), however these have also not been sufficiently validated against actual pilot performance.

The predictive test battery was constructed by first identifying performance variables that should be included. The conceptual model guiding the project indicated that measures reflecting two types of abilities should be included:

- 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.

Thus, the criteria for selecting a variable were that the variable was critical to the task of flying an aircraft and likely to be affected by aging and/or experience. Review of the available taxonomies of piloting behaviors (e. g., Braune & Wickens, 1984; Gerathewohl, 1977; 1978a; 1978b; Imhoff & Levine, 1981) and the aviation human factors and cognitive aging literatures lead to selection of the constructs and processes listed in Table 1. Then the available cognitive skills batteries were reviewed to determine which ones had subtests which mapped most closely to the constructs and processes in Table 1. This analysis lead to the selection of three tests: COGSCREEN, Flitescript and WOMBAT. Each of these tests will be described below. Table 2 provides a summary of the subtests from each of these three measures and, more importantly, how each subtest maps onto the basic piloting processes listed in Table 1.

COGSCREEN. The computerized COGSCREEN test battery was developed over a period of several years by Horst and Kay (1991). The subtests that were selected were based on available task analyses in 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 independent 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 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 the goal of COGSCREEN to reveal cognitive deficits in aviators more readily than would be possible with observation of actual flight under normal conditions. The COGSCREEN tests were not intended to be a comprehensive neuropsychological evaluation battery. Reliability and validity tests for the COGSCREEN battery were not available. However, age norms for experienced pilots were being developed (Horst & Kay, 1991). For this reason, COGSCREEN was selected for inclusion in the predictor battery.

Table 1. Key Piloting Behaviors Most Likely to be 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 subsystems 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 emergency procedures
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 solution (e.g., mechanical failures)

Table 2. Predictor Battery

		COGSCREEN												Flite Scrip	WOMP				
		BACKWARD DIGIT SPAN	VISUAL SEQUENCE COMPARISON	AUDITORY SEQUENCE COMPARISON	MATHEMATICAL PROBLEMS	MANIKIN FIGURES	SYMBOL DIGIT CODING	SYMBOL DIGIT CODING RECALL	MATCHING TO SAMPLE	DIVIDED ATTENTION	PATHFINDER	SHIFTING ATTENTION	DUAL TASK TRACKING	RECOGNITION	TRACKING PERFORMANCE	QUADRANT LOCATION	DIGIT CANCELLING	ERALL SCORE	CHG OVERALL SCORE MID TO END
PERCEPTUAL	AUDITORY PERCEPTION	√		√										√					
	VISUAL PERCEPTION		√			√			√			√			√	√	√	√	
	VISUAL SCANNING						√									√		√	
PSYCHOMOTOR	TRACKING												√		√			√	
ATTENTION	SELECTIVE/FOCUSED	√												√	√	√	√	√	
	DIVIDED									√			√		√			√	
	SWITCHING		√	√			√					√			√	√		√	
	SUSTAINED																	√	√
MEMORY	SHORT TERM AND WORKING MEMORY	√	√	√	√		√		√	√	√		√	√		√	√	√	
	LONG TERM MEMORY				√			√						√					
INFORMATION PROCESSING	SPEED		√	√	√	√			√	√	√	√	√						
	VISUAL SPATIAL					√			√	√			√	√		√		√	
	VERBAL SEQUENTIAL	√	√			√				√	√		√	√		√	√	√	
PROBLEM SOLVING AND DECISION MAKING	DOMAIN INDEPENDENT				√						√	√						√	
	DOMAIN DEPENDENT													√					

WOMBAT. WOMBAT was initially conceived as a pilot selection device for predicting the success of pilot training candidates. WOMBAT measures the ability of the test participant to simultaneously perform several tasks and to determine changing priorities associated with task execution. This requires that the test participant judge the relative worth of a particular action at each moment and dynamically reorder task priorities. To obtain good performance on WOMBAT requires that the test participant develop a strategy for dealing with constantly changing priorities. For this reason, WOMBAT provides a rigorous test of the pilot's ability to attend to varying sources of information and to shift priorities appropriately. In addition, WOMBAT provides a measure of vigilance through a comparison of mid- and end-test segment scores. WOMBAT was selected because (1) it was the only available test that provided information on vigilance and (2) required the test participant to simultaneously attend to several tasks and change priorities dynamically. Furthermore, ongoing WOMBAT data collection efforts will yield normative data on pilots which would provide a standard of comparison for the proposed study.

Flitescript. COGSCREEN and WOMBAT 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 do not measure pilot expertise. Comprehensive evaluation of pilot competence would not be complete without including a measure 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) noted, 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 no well-developed, validated measures that assess this elusive aspect of pilot expertise. The most promising experimental measure of pilot knowledge and judgment was recently developed by Stokes and his colleagues (Stokes, Belger, & Zhang, 1990) at the University of Illinois Institute of Aviation. Flitescript 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. Because the quality of the mental model of the situation may be affected by the availability of situational scripts in long-term memory, it is expected that the performance of novices and experts will differ. That is, 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).

Flitescript appeared to be a promising measure to assess a crucial aspect of pilot knowledge that may be affected by the development of expertise. Therefore, it was included in the predictor battery. The 10-item recognition version of the test was selected since it lends itself to more efficient, objective scoring and yields a response latency score, as well as, an accuracy score.

3.3 Data Collection

Subjects. Forty B727-rated pilots participated in the study. All subjects were volunteers recruited directly from their airlines or the FAA. The initial point-of-contact to the airline and FAA pilots was through Mr. Archie Dillard, National Program Manager, Simulator Engineering, Mike Monroney Aeronautical Center.

All pilots were male and ranged in age from 41 to 71 years (Mean=53.9 years, SD=8.1 years). They were all experienced pilots having a minimum of 5000 hours of flying experience (in any aircraft type). Since not all pilots were active B727 airline pilots, they varied widely in terms of total and recent B727 experience. Subjects' total number of flying hours in the B727 ranged from 4 to 18,000 hours (Mean=5129.8 hours, SD=4685.1 hours). B727 hours in the last 6 months ranged from 0 to 500 hours (Mean=156.25, SD=152.4). B727 hours in the last 30 days ranged from 0 to 225 hours (Mean=28.55, SD=41.04). Thus, the total sample of 40 pilots varied widely in B727 flying experience.

This heterogeneous sample was appropriate to use in two types of analyses. First, this sample's heterogeneity in flying experience allowed for the examination of the relationship between flying experience and performance on the simulator scenarios. It was expected that those pilots with lower levels of B727 experience would perform more poorly in the simulator. This analysis provides some preliminary validity data on the simulator protocol. Second, this larger group of pilots was useful in examining the relationship between pilot age and performance on the predictor tests. For this type of examination, specific type of flying experience (e.g., make and model flown) is unimportant as long as the subject is drawn from the population "professional pilot." The larger sample size lends more power to the analyses looking at the relationship between pilot age and cognitive test performance.

While the 40-pilot heterogeneous sample was useful for the above analyses, for other analyses low levels of total and/or recent B727 experience present a problem. For example, examination of the relationships between predictor test performance and simulator performance and relationships between pilot age and simulator performance are more appropriately done on samples in which all subjects exceed some minimum threshold of B727 experience. Therefore, some analyses were done on a subgroup of 23 subjects.

The 23-pilot subgroup had a very similar age distribution to the total sample (Range 41 to 71 years, Mean=53.0 years, SD=6.87 years), but was more homogenous with respect to B727 flying experience. All 23 of these subjects had at least 500 hours of B727 flying experience. All except 4 subjects had at least 100 hours of B727 flying experience in the last 6 months. Those 4 subjects had lower levels of B727 flying experience (25 to 50 hours during the past 6 months). However, the four subjects were all FAA instructor pilots who had numerous instructor hours in the B727 simulator during the past 6 months. All except one pilot had at least 15 B727 hours in the last 30 days. That pilot had over 250 B727 hours in the last 6 months but had been on a medical leave for the past 30 days. This subgroup of 23 pilots was used in analyses involving the simulator data when the analysis was most appropriately done on experienced B727 pilots.

Experimental procedures. All data collection occurred at the Mike Monroney Aeronautical Center, Oklahoma City. Subjects were recruited through a letter and briefings to the Air Transport Association (ATA) and the Airline Pilots Association with the assistance of Mr. Archie Dillard, simulator manager. The on-site test administrator scheduled the subjects in coordination with Mr. Dillard to assure simulator availability. Subjects were mailed a briefing packet prior to their arrival in Oklahoma City. This packet included information about the purposes and procedures of the study as well as familiarization information about the specific characteristics of the FAA B727 simulator.

Upon arrival in Oklahoma City, subjects were briefed in more detail about study procedures and completed informed consent forms, medical certification record release forms, and a demographic information form that included information related to their occupational history and flying experience. Then, depending on simulator availability and the subjects' arrival time in Oklahoma City, subjects either completed the simulator procedure or the predictor test battery. While the order of testing varied across subjects due to simulator scheduling constraints, detailed records of test scheduling and sequencing were kept for each subject.

The predictor tests were given in two blocks of time. One block (about 2 hours) was for COGSCREEN and Flitescript. The other block (about 3 hours) was for WOMBAT.

The 2 hour simulator flight was preceded with a 30-minute briefing on the configuration of the simulator as well as the role of the observer and first-officer. Subjects piloted the three scenarios and then were debriefed.

Each subject also completed short questionnaires which assessed pilots' perceptions of the difficulty and relevance of each predictor test and the simulator (see Appendix B).

Data preparation procedures. The three PC-based predictor tests recorded each subject's data in a separate file. The on-site test administrator downloaded these files and mailed them to Lehigh for merging and analysis. The pilot observer rating sheets from the simulator protocol were sent to Lehigh for coding and data entry into the computer.

The data collected from each subject included three types of variables: pilot variables, simulator performance variables and predictor test battery variables. Pilot variables were reported by the subject on a demographic questionnaire and included pilot age (in years) and the following flying history variables: total flying hours in all types, total B727 hours, B727 hours in the last 6 months, B727 hours in the last 30 days and B727 simulator hours in the last 30 days.

Simulator performance variables were based on the evaluation of the rater, specifically his rating of the total number of deviations for each maneuver and his subjective evaluation (on the 0 to 100 scale) of each maneuver. Several composite simulator performance variables were created from these ratings. The variable "Total Deviations" was the sum of the deviations for all 23 maneuvers contained in the three scenarios. The variable "Subjective Evaluation" was the mean of the subjective evaluation scores for the 23 maneuvers. The maneuvers were also broken down into three categories: routine, challenging and emergency/abnormal. The 13 routine maneuvers were all relatively low workload highly practiced skills which included normal engine start and taxi, takeoff, turns, nonprecision approach and landing. The 5 challenging maneuvers required a higher level of skill including steep turns, descending turns, precision approach, missed approach and engine-out approach. The 5 emergency/abnormal maneuvers included an aborted take-off, holding pattern entry with an engine flame out, holding pattern with an engine fire, engine-out missed approach and landing with two engines out. The following composite variables were created: total deviations for routine maneuvers, total deviations for challenging maneuvers, total deviations for emergency/abnormal maneuvers, mean subjective evaluation for routine maneuvers, mean subjective evaluation for challenging maneuvers and mean subjective evaluation for emergency/abnormal maneuvers.

The three computer predictor tests yielded a number of variables. The following variables were selected or constructed for use in the analyses.

One variable from Flitescript was utilized in the analyses: percentage of problems answered correctly. Average response latency was not analyzed because there was an extreme range of accuracy scores across subjects making the latency data difficult to interpret.

WOMBAT yields three scores which were used in the analyses: a tracking score, score for the bonus activities, and a total score.

COGSCREEN has 13 subtests which each yield multiple scores. Because of the small number of subjects used in the present experiment it was necessary to reduce the number of variables used in analyses and therefore to construct composite COGSCREEN scores. First, all COGSCREEN variables were examined and variables with an extremely restricted range of subject scores were deleted. For the 22 variables that remained, each subject's score for each variable was converted to a Z score (the subject's score for that variable minus the mean for that variable divided by the standard deviation for that variable). These standardized Z scores were then used to create 3 composite COGSCREEN scores for each subject. The "COGSCREEN Accuracy" variable was the sum of 8 COGSCREEN subtest scores which reflected the number of items a subject had completed correctly on these 8 subtests: Backwards Digit Span, Math Problem, Symbol Digit Coding, Manikin, Divided Attention, Auditory Sequence, Shifting Attention-Discovery and Dual Task Tracking-Previous Number. The "COGSCREEN Speed" variable was the sum of 9 COGSCREEN subtest scores which reflected subjects' average response latency for the following 9 subtests: Math Problems, Visual Sequence Comparison, Matching to Sample, Manikin, Divided Attention, Auditory Sequence, Numerical Pathfinder, Letter Pathfinder and Combined Pathfinder. Note that higher scores for this variable would reflect slower, and therefore poorer performance. The "COGSCREEN Total" variable is the sum of all 22 standardized scores for each subject. This included the 8 accuracy subtest scores, the 9 speed subtest scores, two additional memory recall scores (Symbol Digit Immediate and Delayed Recall), one additional tracking score (Dual Task Tracking-Single), and two additional difference scores (Dual Task Tracking Difference Dual-Single and Divided Attention Difference Dual-Single). In order to create a variable on which higher scores reflect better performance, this variable was created by the addition of all accuracy scores, the addition of the memory scores and then the subtraction of all reaction speed scores, the tracking error score and the two difference scores.

Descriptive statistics (mean, standard deviation and range) for each of the pilot, simulator performance and predictor test variables described above are presented in Appendix C. Data are presented for the total group of 40 subjects and for the subgroup of 23 more experienced B727 pilots.

Data for some variables for some subjects were missing. For example, for several subjects there was a rater omission for the subjective evaluation of one of the simulator maneuvers. Technical problems with the COGSCREEN test also resulted in missing data for several subjects for the Shifting Attention subtest. Finally, several pilots failed to report one or more categories of flight hours on their flying history questionnaire. In all analyses, the following procedure was used to deal with missing data. A subject's data was only included in an analysis if the subject had complete data for the variables included in the analysis. For example, if a correlation between two variables was being calculated, only data from subjects who had complete data for both variables was included. Even though this procedure reduced sample size for some analyses, it was selected rather than another procedure which would have involved estimating and substituting for the missing values. The decision not to use an estimation procedure seemed appropriate given the preliminary nature of this study and the uncertainty of whether the missing values were random in nature. In each of the tables in the next section, the number of subjects with complete data for each variable included in the table is noted. A constant significance level of $p < .05$ was used in all analyses. Depending on changes in sample size (and the resultant changes in degrees of freedom) the critical value associated with this $p < .05$ level varied. For example, the Pearson Product Moment correlation coefficient $r = .32$ is significant at the $p < .05$ level if the correlation is based on data from 38 subjects but is not significant if based on data from fewer subjects.

4.0 RESULTS AND DISCUSSION

4.1 <u>Relationship between Flying Experience and Simulator Performance</u>

The relationship between B727 flying experience and simulator performance was assessed by examination of the Pearson Product Moment correlations presented in Table 3. Correlations were calculated for the total sample which included pilots with a wide range of B727 experience and for the subgroup of pilots who all exceeded a minimum threshold of B727 experience.

Table 3. Correlations between B727 Flying Experience and Simulator Performance

Table 3A. Correlations in the Total Sample

	Total B727 Hours (N=39)	B727 Hours Last 6 Mos (N=36)	B727 Hours Last 30 Days (N=36)	B727 Simulator Hours (N=29)
All Simulator Maneuvers:				
Total Deviations (N=39)	-.19	-.18	-.07	.38*
Subjective Evaluation (N=32)	-.01	.09	.12	.28
Routine Maneuvers:				
Total Deviations (N=39)	-.24	-.24	-.10	-.39*
Subjective Evaluation (N=36)	.17	.25	.18	.25
Challenging Maneuvers:				
Total Deviations (N=40)	-.17	-.10	-.01	-.32
Subjective Evaluation (N=36)	.08	.10	.16	.38*
Emergency/Abnormal Maneuvers:				
Total Deviations (N=40)	-.13	-.15	-.08	-.38*
Subjective Evaluation (N=38)	.15	.23	.16	.19

Table 3B. Correlations in the B727 Experienced Subgroup

	Total B727 Hours (N=22)	B727 Hours Last 6 Mos (N=23)	B727 Hours Last 30 Days (N=23)	B727 Simulator Hours (N= 15)
All Simulator Maneuvers:				
Total Deviations (N=22)	.18	.33	.29	-.62*
Subjective Evaluation (N=19)	-.12	-.02	.07	.32
Routine Maneuvers:				
Total Deviations (N=22)	.06	.19	.24	-.50
Subjective Evaluation (N=20)	-.18	.00	.00	.27
Challenging Maneuvers:				
Total Deviations (N=23)	.17	.45*	.38	.58*
Subjective Evaluation (N=21)	-.17	-.19	.04	.43
Emergency/Abnormal Maneuvers:				
Total Deviations (N=23)	.27	.35	.23	-. 64*
Subjective Evaluation (N=23)	-.14	-.08	-.01	.23

*p<.05.

For the total sample, significant correlations were found between the number of B727 simulator hours the pilots had flown in the past 30 days and the number of deviations they received on the simulator performance measure. A higher number of simulator hours was correlated with fewer deviations. This correlation was significant for two categories of simulator maneuvers (routine and emergency/abnormal) as well as for the composite variable of total deviations across all types of maneuvers. A significant relationship was also found between B727 simulator hours and the raters' subjective evaluation of pilots' performance, but this was only significant for the challenging maneuvers. Pilots who had more recent B727 simulator hours were rated more positively on their performance of the challenging maneuvers.

In the subgroup of more experienced B727 pilots, the pattern of significant correlations between recent B727 simulator hours and simulator performance was very similar. A higher number of B727 simulator hours was correlated with fewer deviations (but in this subgroup only for the challenging and emergency/maneuvers and for the composite total maneuvers variable). In this subgroup of pilots there was also one significant correlation between actual B727 flight hours and simulator performance. Surprisingly, pilots with higher numbers of B727 hours in the last six months had more deviations on their performance of the challenging maneuvers. There is no interpretation for this paradoxical finding that can be empirically examined in the present data. Speculation about its significance may be inappropriate until it is replicated in another sample.

The obtained pattern of correlations are most notable in two respects: the absence of significant relationships between B727 actual flying experience and performance on the simulator measure; and the presence of significant correlations between B727 simulator hours and performance on the simulator measure.

In the total sample of pilots, who were representative of a broad range of B727 flying experience, relationships between amount of actual flying experience and performance on the simulator measure were expected. While several factors may contribute to the failure to obtain such relationships, most importantly it should be noted that B727 hours in this sample were not normally distributed. Pilots seemed to be either very inexperienced (4 pilots had less than 100 total B727 hours) or to have at least 1500 hours. It is likely that there is a "learning curve" for pilot performance such that very inexperienced pilots perform more poorly than pilots who have had some reasonable amount of experience. However, once that threshold of experience has been reached, hours beyond that may not actually correlate with better performance. In other words, pilots with 50 hours of experience may not perform as well as pilots with 1000 hours of experience but pilots with 10,000 hours may not on average perform better than pilots with 2000 hours. In the current sample, the 4 least experienced pilots (who all had less than 100 hours of B727 experience) had an average of 202.5 deviations on the simulator measure. In contrast, the mean number of deviations for the 23 experienced B727 pilots was 104.4 deviations. This suggests that the simulator measure may in fact differentiate between inexperienced and experienced B727 pilots. However, the absence of significant correlations between actual flying hours and simulator performance in experienced B727 pilots is impossible to interpret. There are at least two plausible explanations. Beyond some minimum threshold of B727 experience, increased flying time may in fact be unrelated to improved performance <u>or</u> there may in fact be a relationship but this particular simulator performance measure is not sensitive enough to detect it. Further validation of the simulator performance measure is needed. This examination should include larger numbers of pilots who represent a more continuous distribution of B727 total and recent hours.

The other notable finding from this set of analyses is the presence of significant correlations between recent simulator time and simulator performance. While it is not surprising that those pilots who had the most recent simulator time would perform better in the simulator, the finding should be interpreted with caution due to the small sample size. Only 29 of the 40 pilots completed this measure of flying experience, data for the other pilots was missing for this variable. Of the 29 pilots who completed this item, 16 reported 0 B727 simulator hours in the last month. Also, it should be noted that at least 10% of the present sample of pilots were in training or instructor positions with the FAA or their air carrier and had a great deal of simulator experience, not simply the reported hours they had actually flown themselves but also many unreported observation hours in the simulator. For these reasons (small sample size, nonnormal distribution of simulator hours, and inclusion of instructor pilots) the significant correlations between recent B727 simulator hours and performance on the simulator performance measure should not be over interpreted. In further efforts to validate the simulator performance measure, data from instructor pilots should be gathered in sufficient numbers to allow for separate analyses.

4.2 Relationship between the Predictor Tests and Simulator Performance

Pearson Product Moment correlations between the predictor test variables and the simulator performance variables are presented in Table 4. For completeness, the correlations are presented for the total sample of pilots and for the subgroup of B727 experienced pilots. However, examination of the relationship between the cognitive predictor variables and simulator performance is most appropriate when pilots exceed some minimum threshold of B727 experience. The critical question is whether the predictors can discriminate the varying levels of simulator performance in an experienced pilot sample. Therefore, the results for the more homogeneous B727 experienced subgroup are focused on in this discussion.

As shown in Table 4B, none of the Flitescript or WOMBAT variables were significantly correlated with any of the simulator performance variables. The COGSCREEN total composite variable (which reflected both speed and accuracy subtests as well as memory and tracking subtests) was correlated with the raters' subjective evaluation of pilots' performance of the emergency/abnormal maneuvers.

These results suggest that COGSCREEN does have potential to discriminate the simulated flying performance of experienced B727 pilots, particularly when pilots are required to perform under unusual, high workload emergency conditions. It is possible that COGSCREEN might have shown even stronger relationships to simulator performance if sample size would have been large enough to allow for examination of relationships between specific COGSCREEN subtests (rather than the composite variables) and specific components of flying performance. For example, COGSCREEN tracking performance may have been

Table 4. Correlations between the Predictor Tests and Simulator Performance

Table 4A. Correlations in the Total Sample

	Flitescript	WOMBAT			COGSCREEN		
	% Correct (N=40)	Track (N=40)	Bonus (N=40)	Total (N=40)	Accur. (N=34)	Speed (N=39)	Total (N=33)
All Simulator Maneuvers:							
Total Deviations (N=39)	-.12	.04	.00	.03	-.25	.00	-.05
Subjective Evaluation (N=32)	.15	.08	.19	.15	.19	-.08	.14
Routine Maneuvers:							
Total Deviations (N=39)	-.11	.02	-.11	-.04	-.17	.02	-.02
Subjective Evaluation (N=36)	.26	-.02	.05	.02	.27	-.03	.04
Challenging Maneuvers:							
Total Deviations (N=40)	-.12	-.04	.01	-.02	-.31	.05	-.09
Subjective Evaluation (N=36)	.10	.04	.14	.09	.07	-.03	.02
Emergency/Abnormal Maneuvers:							
Total Deviations (N=40)	-.11	.14	.13	.16	-.21	-.06	-.03
Subjective Evaluation (N=38)	.24	-.07	-.02	-.06	.34*	-.04	.10

Table 4B. Correlations in the B727 Experienced Subgroup

	Flitescript	WOMBAT			COGSCREEN		
	% Correct (N=40)	Track (N=40)	Bonus (N=40)	Total (N=40)	Accur. (N=34)	Speed (N=39)	Total (N=33)
All Simulator Maneuvers:							
Total Deviations (N=22)	.14	-.18	-.15	-.19	-.26	.20	-.37
Subjective Evaluation (N= 19)	.07	.30	.32	.35	.38	-.32	.44
Routine Maneuvers:							
Total Deviations (N=22)	.03	-.24	-.25	-.27	-.24	.31	-.39
Subjective Evaluation (N=20)	.06	.27	.29	.32	.34	-.29	.38
Challenging Maneuvers:							
Total Deviations (N=23)	.19	-.19	-.11	-.10	-.20	.14	-.27
Subjective Evaluation (N=21)	.02	.36	.27	.36	.35	-.35	.42
Emergency/Abnormal Maneuvers:							
Total Deviations (N=23)	.10	-.02	.06	.02	-.15	.07	-.28
Subjective Evaluation (N=23)	.11	.24	.21	.26	.42	-.39	.48*

*$p<.05$.

related to performance of challenging maneuvers such as steep turns or descending turns. This is the type of hypothesis that should be explored in future research.

While the results provide no evidence of the ability of Flitescript or WOMBAT to predict simulator performance, it should be remembered that the power to detect such relationships was low due to the small sample size.

4.3 Relationship Among the Predictor Tests

One goal of the present study was to reduce redundancy and construct the most efficient battery of predictor tests. Therefore correlations between each of the predictor test variables were calculated and the results are presented in Table 5. For completeness, correlations were calculated in the total sample and the B727 experienced subgroup. However, since simulator performance did not enter into these correlations it is appropriate to emphasize the correlations in the total sample. Although it can be noted that a very similar pattern of correlations occurred for the B727 experienced subgroup of pilots. Correlations between a subtest (such as WOMBAT bonus or COGSCREEN accuracy) and total test performance (such as WOMBAT total or COGSCREEN total) are not presented since those two types of variables are not independent.

Flitescript was not significantly correlated with any WOMBAT or COGSCREEN variables. This was not unexpected since Flitescript was explicitly selected to be a measure of pilots' domain-dependent aviation knowledge and judgment, not a measure of the domain-independent cognitive skills assessed by WOMBAT and COGSCREEN. However, if Flitescript is a measure of domain-dependent aviation knowledge, it is reasonable to hypothesize that it would correlate with pilots' level of aviation experience. In the present study, Flitescript was not found to be correlated with any measure of flying experience, including total hours in all types ($r=-.10$). This lack of significant relationship should be interpreted with caution due to the small sample size and nonnormal distribution of pilot hours. Although, as will be discussed in section 4.5, there is some indication from the present data that Flitescript may need to be revised if this test is to be a highly correlated measure of pilot experience.

Within WOMBAT, there was a significant positive correlation between scores on the tracking task and scores on the bonus tasks. This may at least partially be due to differences in pilots' experience with this type of "video game" and their resultant comfort and interest in the WOMBAT tasks. Pilots who had more experience with these types of video games performed better on both WOMBAT's tracking ($r=.32$) and bonus ($r=.32$) tasks. Experience with this type of game might facilitate subjects performance on the tracking task and thus allow more effort to be placed on the bonus tasks resulting in higher scores on both.

COGSCREEN performance on the accuracy measures and speed measures was found to be negatively correlated. The faster a pilot performed on the subtests the lower his accuracy was likely to be. This is not an unusual finding and may simply reflect individual differences in preference for and emphasis on accurate or fast performance. An important question is whether

Table 5. Correlations among the Predictor Tests

Table 5A. Correlations in the Total Sample

	Flitescript % Correct (N=40)	WOMBAT Track (N=40)	Bonus (N=40)	Total (N=40)	COGSCREEN Accur. (N=34)	Speed (N=39)	Total (N=33)
WOMBAT:							
Tracking	.38						
Bonus	.28	.41*					
Total	.31						
COGSCREEN:							
Accuracy	.17	.31	.44*	.43*			
Speed	-.28	-.21.	-.32*	-.30	-.47*		
Total	.05	.32	.39*	.42*			

Table 5B. Correlations in the B727 Experienced Subgroup

	Flitescript % Correct (N=40)	WOMBAT Track (N=40)	Bonus (N=40)	Total (N=40)	COGSCREEN Accur. (N=34)	Speed (N=39)	Total (N=33)
WOMBAT:							
Tracking	.38						
Bonus	.35	.57*					
Total	.41						
COGSCREEN:							
Accuracy	.08	.30	.51*	.45*			
Speed	-.31	-.15	-.51*	-.36	-.57*		
Total	.03	.24	.53*	.42			

*p<.05

this speed accuracy tradeoff is related to pilots' age. This will be considered in the next section.

Many significant correlations were found between performance on WOMBAT and performance on COGSCREEN. WOMBAT bonus score was significantly correlated with all three COGSCREEN variables. Pilots who performed better on the WOMBAT bonus tasks also performed more accurately and faster on the COGSCREEN subtests. Total WOMBAT score was also correlated with COGSCREEN accuracy and total scores. Relationships between these two measures of domain-independent cognitive abilities is consistent with the interpretation that both measures assess similar underlying cognitive abilities. Decisions concerning which measure should be included in subsequent research on pilot performance should be based on practical issues such as ease of administration and scoring but most importantly on which measure correlates most highly with pilot performance in the simulator. In the present study, COGSCREEN correlated with simulator performance while WOMBAT did not.

4.4 Relationships between Pilot Age and Flying Experience, Simulator Performance, and the Predictor Tests

Table 6 presents correlations between pilots' age and their B727 flying experience, simulator performance and predictor test performance. Correlations in the total sample and the B727 experienced subgroup are presented. Again, correlations which involve any aspect of simulator performance are best examined in the experienced subgroup.

Pilot age was found to be significantly correlated with simulator performance in the experienced subgroup. Older pilots were given lower subjective evaluation ratings on all three types of maneuvers. Since there were no significant correlations between age and B727 flying experience, this finding of poorer simulator performance in the older pilots does not seem to be related to age differences in B727 experience. It is also noteworthy that age was correlated with the subjective ratings but not the deviation scoring. It is clear from the current data that the two scoring systems are related but not identical. The total number of deviations for all 23 maneuvers and the mean of the subjective evaluations for those maneuvers are significantly but not perfectly correlated ($r=.61$). Further development and validation of the simulator measure should explore the properties of these two scoring systems.

One possible interpretation of this pattern of results is that deviation scoring and subjective evaluation rating are not equally sensitive in detecting actual age effects in simulated flying performance. However, it may also be possible that the evaluation rating system (since it is more global and subjective than the deviation scoring system) is more vulnerable to potential "age bias" on the part of the raters. It is not possible to test these alternate explanations with the current data set. However, this preliminary finding that increasing pilot age was correlated with poorer simulator performance is noteworthy and should be examined further in a larger sample with an appropriate age distribution of pilots.

Pilot age was also significantly correlated with performance on the predictor tests. Here the critical question is whether there is a relationship between chronological age and performance on these cognitive tests in "professional pilots" (not just B727 pilots). Therefore,

Table 6. Correlations Between Pilot Age and Flying Experience, Simulator Performance and the Predictor Tests

	Pilot Age			
	Total Sample		B727 Experienced Subgroup	
	r	(N)	r	(N)
Total B727 Hours	.19	(39)	.15	(22)
B727 Hours Last 6 Months	-.04	(36)	-.09	(23)
B727 Hours Last 30 Days	-.01	(36)	-.02	(23)
B727 Simulator Hours	-.10	(29)	-.17	(15)
All Simulator Maneuvers:				
Total Deviations	-.01	(39)	.11	(22)
Subjective Evaluation	-.21	(32)	-.49*	(19)
Routine Maneuvers:				
Total Deviations	.01	(39)	.23	(22)
Subjective Evaluation	-.09	(36)	-.48*	(20)
Challenging Maneuvers:				
Total Deviations	.02	(40)	.08	(23)
Subjective Evaluation	-.06	(36)	-.51*	(21)
Emergency/Abnormal Maneuvers:				
Total Deviations	-.05	(40)	-.02	(23)
Subjective Evaluation	-.14	(38)	-.48*	(23)
Flitescript:				
% Correct	-.35*	(40)	-.37	(23)
WOMBAT:				
Tracking	-.43*	(40)	-.38	(23)
Bonus	-.18	(40)	-.48*	(23)
Total	-.39*	(40)	-.48*	(23)
COGSCREEN:				
Accuracy	-.38*	(34)	-.47*	(19)
Speed	.51*	(39)	.71*	(23)
Total	-.46*	(33)	-.54*	(19)

*p<.05

it is appropriate to focus on the total group and take advantage of the increased power this larger sample provides. In this total sample, scores on all three predictor tests are correlated with pilot age. Older pilots completed fewer Flitescript problems correctly; performed more poorly on WOMBAT (particularly on the tracking task); and, performed more poorly on COGSCREEN (with lower levels of accuracy and longer reaction times). These findings were expected since the batteries, particularly WOMBAT and COGSCREEN, were explicitly selected for inclusion in the predictor battery because they assessed aviation-relevant abilities that prior literature suggested were most likely to be affected by aging.

Considering the pattern of significant correlations between pilot age and the simulator measure, between pilot age and the predictor tests and between the predictor tests and the simulator an interesting pattern of inter-correlations appears. Pilot age is correlated with COGSCREEN performance and both pilot age and COGSCREEN are correlated with simulator performance. Here the critical question becomes which correlated variable, pilot age or COGSCREEN, is a stronger predictor of simulator performance. In other words, is pilot age still significantly correlated with simulator performance when the contribution of COGSCREEN performance has been removed and/or is COGSCREEN still significantly correlated with simulator performance when the contribution of pilot age has been removed. This pattern of inter-correlations lends itself to partial correlation analyses. This type of analysis requires that only subjects with complete data for all three inter-correlated variables be included in the analysis. In the present study there were 19 B727 experienced pilots who had complete data for these three variables: pilot age, the mean subjective evaluation for all simulator maneuvers variable and the COGSCREEN total variable. Correlations among these three variables (for these 19 pilots) were as follows: the correlation between age and COGSCREEN was -.68; the correlation between age and the simulator was -.58; and, the correlation between COGSCREEN and the simulator was .54. Thus, age and COGSCREEN were significantly correlated with each other and both were significantly correlated with the simulator. The partial correlation between age and the simulator with the contribution of COGSCREEN partialed out was -.34, which was no longer significant. The partial correlation between COGSCREEN and the simulator with the contribution of pilot age partialed out was .24, which was also no longer significant. Thus, while this analysis was appropriate for investigating this pattern of inter-correlations, the results for this small sample of pilots do not lead to clear interpretation. All that can be concluded at this point is that there may be something unique to the intercorrelation between age and COGSCREEN performance that contributes to simulator performance above and beyond the separate contribution of each variable. Further research examining pilot age, COGSCREEN performance and simulator performance is needed.

4.5 Pilots' Perceptions of the Simulator and Predictor Tests

Since one goal of this project was to examine the feasibility of a comprehensive, quantitative approach to pilot performance assessment, pilot acceptance of these types of measures was an issue of interest. Therefore, short questionnaires assessing pilots' perceptions of the three predictor tests and the simulator performance measure were administered. These questionnaires, and descriptive statistics summarizing pilots' responses, are presented in Appendix B.

Pilot acceptance was clearly highest for the simulator measure. On average, pilots perceived the simulator measure to be "comprehensive", a "good measure of flying ability" and a "valid method for screening unsafe pilots." Most pilots also reported that they "would not object to taking this test on a regular basis."

Pilot perceptions of the domain-independent cognitive skills predictor tests, COGSCREEN and WOMBAT, tended to be less positive. On average, pilots were not sure that these tests tap the "cognitive abilities that one would expect to find in a safe pilot" and they did not tend to feel these tests would be "a reasonably valid method for screening unsafe pilots."

Flitescript was included in the predictor battery because it was hypothesized to be a measure of domain-dependent aviation knowledge and judgment which should be related to amount of flying experience. The pilots did not perceive that Flitescript was "a good measure of the experience level of a pilot." Further, they tended to feel that Flitescript was not "a reasonably valid method for screening unsafe pilots." Several pilots spontaneously reported that they felt the test was more suited for air traffic controllers than pilots.

5.0 CONCLUSIONS

The long-term aim of the research presented in this report is to increase understanding about the relationships among pilot age, experience, and performance. This information is critical to making informed decisions about the Age 60 Rule. The specific goals of the present project were to:

- Develop a criterion measure of complex pilot performance in a simulator that was objective and quantifiable.
- Develop a test battery of component skills and abilities that were assumed to contribute to complex performance. The test battery measurements served as predictor variables of simulator performance.
- Collect data on the predictor variables and the simulator criterion measure within a group of 40 pilots varying in age.

5.1 Conclusions about the Simulator Performance Measure

The Lehigh/Hilton Systems research team worked with FAA personnel (including simulator experts, instructors and check pilots) to develop a simulator assessment procedure that was more objective, quantifiable, and definitive than the typical proficiency "check-ride." Three simulator scenarios were developed that included routine maneuvers, abnormal situations and emergency flight conditions. These maneuvers were selected to be realistic and credible, as well as challenging. The pilots who participated in this study strongly agreed that the simulator scenarios were "comprehensive" and "a good measure of flying ability." Thus, the simulator scenarios appear to have good face validity with the pilots.

Two types of quantitative scoring systems were developed for the simulator performance measure. Evaluators were trained to observe a pilot's performance of each maneuver and to assign a subjective overall rating for each maneuver using a 0 to 100 scale. Evaluators were also trained to use a deviation scoring system which involved using a detailed maneuver scoring sheet to record the number of deviations from error free performance for each maneuver. The results of the present study demonstrate that these two types of scoring systems yield similar but not identical evaluations of pilot performance. The total number of deviations for all 23 maneuvers and the mean of the subjective evaluations for those maneuvers are significantly but not perfectly correlated ($r=.61$). Further development and validation of the simulator measure should explore the properties of these two scoring systems.

This preliminary effort to develop a procedure for assessing pilots' performance in a simulator suggests that objective, quantifiable assessments are possible. Further validation of this particular simulator performance measure (including the scoring systems and selection of maneuvers) is needed. The most powerful validation approach would involve correlating pilots' performance on this simulator measure with other measures of their flying performance (e.g., supervisor or instructor ratings or flying history such as previous involvement in accidents or incidents). However, this stringent type of validation may not be possible. Efforts could also be made to further explore relationships between amount of flying experience and performance on the simulator measure using a larger sample of pilots with a more normal distribution of B727 flight hours. Formal demonstration of evaluators' ability to use the scoring systems consistently (high inter-rater reliability) will also be required. Finally, the generalizability of this performance assessment measure to other aircraft makes and models will need to be explored.

5.2 Conclusions about the Predictor Battery

The second task of the project was to construct a predictive test battery which could be validated against the simulator criterion measure. Two measures of domain-independent cognitive skills (COGSCREEN and WOMBAT) and one measure of domain-dependent aviation knowledge and judgment (Flitescript) were selected for inclusion in this predictor battery.

One goal of the study was to examine the relationship between the predictor tests and performance in the simulator. Performance on Flitescript and WOMBAT was not found to be significantly correlated with any of the simulator performance variables. The COGSCREEN total composite variable (which reflected both speed and accuracy subtests, as well as memory and tracking subtests) was correlated with the raters' subjective evaluation of pilots' performance of the emergency/abnormal maneuvers.

These results suggest that COGSCREEN does have potential to discriminate the simulated flying performance of experienced B727 pilots, particularly when pilots are required to perform under unusual, high workload emergency conditions. It is possible that COGSCREEN might have shown even stronger relationships to simulator performance if sample size would have been large enough to allow for examination of relationships between specific COGSCREEN subtests (rather than the composite variables) and specific components of flying performance. For example, COGSCREEN tracking performance may have been related to performance of challenging maneuvers such as steep turns or descending turns. This is the type of hypothesis that should be explored in future research.

Another goal of the study was to examine the relationships among the predictors to reduce redundancy and provide the most concise battery of predictor tests. Flitescript was not found to be significantly correlated with any WOMBAT or COGSCREEN variables. This was not unexpected since Flitescript was explicitly selected to be a measure of pilots' domain-dependent aviation knowledge and judgment, not a measure of the domain-independent cognitive skills assessed by WOMBAT and COGSCREEN. Many significant correlations were found between performance on WOMBAT and performance on COGSCREEN. WOMBAT bonus score was significantly correlated with all three COGSCREEN variables. Pilots who performed better on the WOMBAT bonus tasks also performed more accurately and faster on the COGSCREEN subtests. Total WOMBAT score was also correlated with COGSCREEN accuracy and total scores. Relationships between these two measures of domain-independent cognitive abilities is consistent with the interpretation that both measures assess similar underlying cognitive abilities.

Decisions concerning which predictor test measures to include in subsequent research on pilot performance should be based on practical issues such as ease of administration and scoring but most importantly based on which measures correlate most highly with pilot performance in the simulator. In the present study, COGSCREEN correlated with simulator performance while WOMBAT and Flitescript did not. It is unfortunate that Flitescript did not correlate with simulator performance since it was the only measure in the predictor battery which was hypothesized to reflect domain-dependent aviation knowledge and judgment. Efforts should be made to refine or develop and test measures that are reflective of pilots' aviation-relevant experience and judgment.

5.3 Conclusions about Pilot Age and Performance on the Simulator Measure and the Predictor Tests

A final goal of this project was to conduct a preliminary examination of the relationship between pilot age and performance on the simulator measure and predictor battery. Prior to discussion of these results, it should be reiterated that the sample size in this study was small, the age range was broad (age 41 to 71 years) and unevenly distributed, and that pilots older than the typical air carrier pilot (who now retires at or before age 60) participated.

Pilot age was found to be related to simulator performance. Increasing pilot age was correlated with poorer evaluator rating of simulator performance. This finding is noteworthy but should not be overinterpreted until it is replicated in a larger sample with an appropriate age distribution of pilots. Interpretation of this finding may also be clarified by further efforts to validate this particular simulator performance measure and the two scoring systems.

Pilot age was also found to be significantly correlated with performance on the predictor tests. Older pilots completed fewer Flitescript problems correctly; performed more poorly on WOMBAT (particularly on the tracking task); and, performed more poorly on COGSCREEN (with lower levels of accuracy and longer reaction times). These findings were expected since the batteries, particularly WOMBAT and COGSCREEN, were explicitly selected for inclusion in the predictor battery because they assessed aviation-relevant abilities that prior literature suggested were most likely to be affected by aging.

Considering the significant correlations between pilot age and the simulator measure, between pilot age and the predictor tests and between the predictor tests and the simulator that emerged in this study, an interesting pattern of inter-correlations appears. Pilot age was correlated with COGSCREEN performance and both pilot age and COGSCREEN were correlated with simulator performance. Here the critical question becomes which correlated variable, pilot age or COGSCREEN, is a stronger predictor of simulator performance. In other words, is pilot age still significantly correlated with simulator performance when the contribution of COGSCREEN performance has been removed and/or is COGSCREEN still significantly correlated with simulator performance when the contribution of pilot age has been removed. These questions can potentially be addressed through partial correlation analyses. However, the results for this small sample of pilots did not lead to clear interpretation. All that can be concluded at this point is that there may be something unique to the inter-correlation between age and COGSCREEN performance that contributes to simulator performance above and beyond the separate contribution of each variable. Further research examining pilot age, COGSCREEN performance and simulator performance is needed.

REFERENCES

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods.* Monterey, CA: Brooks/Cole.

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.

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

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.

Carretta, T. R. (1987). *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.

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.

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.

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* (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/MFA01).

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.

Hyland, D. T., Kay, E. J., Deimler, J., & Gurman, E. (1991). *Airline pilot age and performance: A review of the scientific literature* (Contract No. DTFA-02-90-90125). Oklahoma City, OK: Civil Aeromedical Institute, Federal Aviation Adminstration.

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.

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.

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.

McCall, R. B. (1970). *Fundamental Statistics for Psychology*, Harcourt-Brace & World, NY.

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).

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

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.

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

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. (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.

Schaie, K. W. & Baltes, P. B. (1975). On sequential strategies and developmental research, *Human Development, 18*, 384-390.

Stokes A. F., Banich M. T., & Elledge, V. (1988). *Cognitive function evaluation in the medical certification of airmen* (Final Technical Report 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.

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

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

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

APPENDIX A

The Simulator Performance Measure

Control Number ______

SCENARIO I: Preflight checklist, before start checklist, engine start, taxi, takeoffs, climbs, turns, climbing turns, descending turns, steep turns, level flight, tracking, precision approach, and landing.

The grading criterion for each of these scenarios is listed on the following pages in the order flown. The performance will be recorded as the maneuvers are flown. At the completion of each maneuver the evaluator will enter his numerical assessment. At the end of the scenario the evaluator will record his assessment of the entire flight.

Pilot instructions:

Aircraft weight: 150000 lbs.

ATIS: OKC weather 40 SCT 4 70/56 270008 992 Active RY35R

Flight Information: Perform a basic preflight cockpit check, using the simulator checklist. Start engines and taxi to the runway assigned; perform necessary checklists. Takeoff RY35R and climb to 5000 ft; level off and set aircraft up for level flight at 250 kts. When in level flight and at 250 kts, make a 180 degree turn to the right and roll out on heading 170 degrees, maintaining speed and altitude. Next, turn left to heading 350 degrees while climbing to 6000, maintaining 250 kts. When at 6000 ft and on heading 350 degrees, make a steep turn of 360 degrees using 45 degrees of bank to a heading of 350 degrees, maintaining speed and altitude. Upon completion of steep turn, make a descending left/right turn to 5000 ft and roll out on heading 170 degrees. At this point pick up an IFR clearance for ILS RY35R APCH via direct GALLY; descend and maintain 4000 ft. At GALLY aircraft cleared for approach and landing. Weather at OKC is now E7 BKN 50 OVC 2 70/56300012 992.

JULY 1, 1992

Control Number____________

SCENARIO II: Before engine start checklist, engine start, IFR clearance, aborted takeoff, takeoff, weather, holding pattern, flameout, VOR nonprecision approach, missed approach, VFR landing.

The grading criterion for each of these scenarios is listed on the following pages in the order flown. The performance will be recorded as the maneuvers are flown. At the completion of each maneuver the evaluator will enter his numerical assessment. At the end of the scenario the evaluator will record his assessment of entire flight.

Pilot instruction:

Aircraft weight: 150000 lbs.

ATIS: OKC weather thin obscuration (-X) 45 BKN 3/4 V1 GF 70/67 light and variable 992 RY___. DFW weather M30 3 R- 170010 990.

Flight Information: IFR clearance - N727 CLRD DFW AIRPORT VIA FLY RY HDG/4000 THEN TURN RIGHT HDG 180 DEGREES UNTIL JOINING J21 THEN AS FILED; CLIMB AND MAINTAIN 8000 FT; EFC FL230 10 MIN AFTER DEPT; CONTACT DEPT CONTROL 124.6; SQUAWK 3400. After aborted takeoff due to engine fire before V1, reissue same route clearance; climb and maintain 6000 ft. After aircraft is airborne, a new clearance is issued due to weather and traffic on DFW. IFR clearance - N727 CLRD DIRECT IRW VORTAC; MAINTAIN 5000; HOLD SOUTH ON 180 RADIAL EFC________(+10 min). After aircraft is established in the holding pattern, number 3 engine will flameout as aircraft turns inbound on second pattern (outbound end). New clearance - N727 CLRD VOR RY17L APCH. Restate the weather and include a NOTAM - All ILS systems at OKC are inoperative. At MDA, the runway is not visible to the pilot. The pilot should execute a missed approach [weather below minimums/or runway not visible]. After a positive rate of climb is established, the weather will change and immediately become CLR 20 70/60 310015 992. N727 will be issued the new weather INCORRECTLY and cleared VIA VECTORS for a visual landing RY30.

FLIGHT PLAN:

OKC-DIRECT-IRW-J21-ADM R-110 BENCH-BUJ6-DFW FL230

Revised page 7/20/92

Control Number ________ ___

SCENARIO III: Before engine start checklist, engine start, IFR clearance, weather, takeoff, new IFR clearance, engine inoperative as entry made into holding pattern (engine fire), runway/approach selection, back course ILS RY35L, two engine missed approach, one engine operation to VFR landing.

The grading criterion for each of these scenarios is listed on the following pages in the order flown. The performance will be recorded as the maneuvers are flown. At the completion of each maneuver the evaluator will enter his numerical assessment. At the end of the scenario the evaluator will record his assessment of the entire flight.

Pilot instructions:

Aircraft weight: 140000 lbs.

ATIS: OKC WX M9 OVC 4 70/67 360012 992 ACTIVE RY35R.

Flight information: IFR clearance - N727 CLRD TO HOU AIRPORT VIA FLY RY HDG UNTIL REACHING 4000; THEN TURN RIGHT HDG 180 UNTIL JOINING J21; THEN AS FILED; CLIMB AND MAINTAIN 8000 FT; EXPECT FL270 10 MIN AFTER DEPT; CONTACT DEPARTURE CONTROL 124.6; SQUAWK 3400. After departure as aircraft climbs at 3500 ft, a new clearance is issued due to aircraft in an emergency making an ILS RY35 APCH. N727 MAINTAIN 4000; TURN LEFT DIRECT TO IRW VORTAC; HOLD SOUTH ON 180 RADIAL; EXPECT FURTHER CLEARANCE (+10 MIN)_____. After aircraft enters holding pattern, a fire is created in number 3 engine second time over fix and turning outbound. After fire is under control, give info that the emergency aircraft is disabled on RY35R. Issue radar vectors to the back course ILS RY35R and a clearance for BC RY35L APCH. A missed approach will be forced at MDA (VEHICLE ON RUNWAY). During the missed approach, a second engine loss will occur (5 degs flaps and 170 kts). The weather will improve to CLR 20 70/50 120010 992. N727 will be cleared for landing on any runway with only one engine.

FLIGHT PLAN:

OKC-IRW-J21-DFW-787-TNV (NAVASOTA) STRUK7 HOU

Revised page 7/20/92

Control Number____________

MANEUVER: PREFLIGHT/TAXI (CATEGORIZATION: ROUTINE)

FUNCTION	CRITERION	SCORE	REMARKS
Preflight checklist	Set correctly	A	
	Omitted	B	
	Set incorrectly	C	
Before start checklist	Set correctly	A	
	Omitted	B	
	Set incorrectly	C	
Taxi checklist	Checked correctly	A	
	Omitted/moving	B	
	Checked incorrectly	C	
Crew briefing	Briefing	A	
	Incomplete briefing	B	
	No briefing	C	
Runway alignment	Centerline	A	
	1/4 L/R centerline	B	
	+1/4 L/R centerline	C	

EVALUATOR ASSESSMENT

______

BELOW AVERAGE				AVERAGE				ABOVE AVERAGE				0 ERRORS	
	50		60	70		75		80	85	90	95		100
40		55	65	68	72		76	78	84	88	94	98	

NEXT MANEUVER: **CLEARED FOR TAKEOFF; FLY RY HEADING UNTIL 5000 FT; IAS 250 KTS.**

Control Number____________

MANEUVER: TAKEOFF (CATEGORIZATION: ROUTINE)
CLRD FOR TAKEOFF; FLY RY HEADING UNTIL 5000 FT; SPEED 250 KTS. AT 12 - 15 MILES DME FROM IRW VORTAC, ISSUE INSTRUCTIONS FOR NEXT MANEUVER.

FUNCTION	CRITERION	SCORE	REMARKS
Thrust managment	2 of 2 steps	A	
	1 of 2 steps	B	
	Unorthodox	C	
Direction control	Rudder/Tiller	A	
	Rudder	B	
	Tiller	C	
Cockpit communication	80 kts, V1Vr	A	
	1 of 2 speeds	B	
	No speed calls	C	
Rotation pitch	15 deg	A	
	-15 deg	B	
	+15 deg	C	
Speed at rotation	V1Vr	A	
	V1Vr +/-5 kts	B	
	V1Vr +/-10 kts	C	
	V1Vr +/-20 kts	D	
Heading	+/-5 deg	A	
	+/-10 deg	B	
	+/-15 deg	C	
Cockpit communication	Gear up call	A	
	Late gear up call	B	
	Omitted call	C	
Speed (attitude)	(V2+10)+/-5 kts	A	
	(V2+10)+10 kts	B	
	(V2+10)-10 kts	C	

Control Number________

MANEUVER: TAKEOFF (Cont'd)

FUNCTION	CRITERION	SCORE	REMARKS
After takeoff checklist 1000 ft cleanup	Correct Omitted Incorrect	A B C	
Flaps	Calls/steps (3) 2 of 3 steps 1 of 3 steps	A B C	
Heading	+/-0 deg +/-5 deg +/-10 deg +/-15 deg	A B C D	
Altitude	+/-0 ft +/-50 ft +/-100 ft +100 ft	A B C D	
Major Deviation	Alt beyond 500, loss of control, off ry, etc.	G	
EVALUATOR ASSESSMENT		______	

BELOW AVERAGE			AVERAGE			ABOVE AVERAGE			0 ERRORS	
	50	60	70		75	80	85	90	95	100
40	55	65	68	72	76	78	84	88	94	98

NEXT MANEUVER: (AT 12 - 15 MILES NORTH) MAKE A 180 DEGREE TURN TO THE RIGHT; ROLL OUT HEADING 170 DEGREES; MAINTAIN IAS 250 KTS AND 5000 FT.

Control Number____________

MANEUVER: TURNS (CATEGORIZATION: ROUTINE)
MAKE A 180 DEGREE TURN TO THE RIGHT; ROLL OUT HEADING 170 DEGREES. MAINTAIN IAS 250 KTS AND 5000 FT.

FUNCTION	CRITERION	SCORE	REMARKS
Turns	30 deg bank	A	
	30 deg bank +/-5 degs	B	
	30 deg bank +/-10 degs	C	
	30 deg bank +/-15 degs	D	
Altitude	+/-0 ft	A	
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Speed	+/-0 kts	A	End of turn
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Heading (roll out)	+/-0 deg	A	
	+/-5 deg	B	
	+/-10 deg	C	
	+/-15 deg	D	
Major Deviation	Alt beyond 500 ft, speed +250 kts, etc.	E	

EVALUATOR ASSESSMENT

______

BELOW AVERAGE			AVERAGE			ABOVE AVERAGE			0 ERRORS	
	50	60	70		75	80	85	90	95	100
40	55	65	68	72	76	78	84	88	94	98

<u>NEXT MANEUVER</u>: MAKE A 180 CLIMBING TURN TO THE LEFT AT 1000 FT PER MIN; ROLL OUT HEADING 350 DEGREES AND 6000 FT; MAINTAIN IAS 250 KTS.

Control Number__________

MANEUVER: CLIMBING 180 DEGREE TURNS (CATEGORIZATION: ROUTINE)
MAKE A 180 CLIMBING TURN TO THE LEFT AT 1000 FEET PER MIN; ROLL OUT HEADING 350 DEGREES AT 6000 FEET AND MAINTAIN IAS 250 KTS.

FUNCTION	CRITERION	SCORE	REMARKS
Turns	30 deg bank	A	
	30 deg bank +/-5 degs	B	
	30 deg bank +/-10 degs	C	
	30 deg bank +/-15 degs	D	
Climb rate (1000 ft per min)	+/-100 ft	A	
	+/-200 ft	B	
	+/-300 ft	C	
Speed	+/-0 kts	A	
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Heading (roll out)	+/-0 deg	A	
	+/-5 deg	B	
	+/-10 deg	C	
Altitude	+/-0 ft	A	
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Major Deviation	Loss of control, etc.	E	

EVALUATOR ASSESSMENT ______

BELOW AVERAGE				AVERAGE				ABOVE AVERAGE				0 ERRORS	
	50	60		70		75		80	85	90	95		100
40		55	65	68	72		76	78	84	88	94	98	

<u>NEXT MANEUVER</u>: MAKE A 360 DEGREE STEEP TURN TO THE RIGHT (45 DEGREES OF BANK); ROLL OUT HEADING 350 DEGREES; MAINTAIN IAS 250 KTS AND 6000 FT.

Control Number____________

MANEUVER: STEEP TURNS (CATEGORIZATION: ROUTINE)
MAKE A 360 DEGREE STEEP TURN (45 DEGREES OF BANK) TO THE RIGHT; ROLL OUT HEADING 350 DEGREES; MAINTAIN 250 KTS AND 6000 FEET.

FUNCTION	CRITERION	SCORE	REMARKS
Angle of bank	45 deg	A	Ck at 90 degs
	45 deg +/-5 deg	B	
	45 deg +/-10 deg	C	
	45 deg +/-15 deg	D	
Altitude	+/-0 ft	A	Ck at 180 degs
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Speed	+/-0 kts	A	Ck at 180 degs
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Angle of bank	45 deg	A	Ck at 270 degs
	45 deg +/-5 deg	B	
	45 deg +/-10 deg	C	
	45 deg +/-15 deg	D	
Altitude	+/-0 ft	A	Ck at roll out
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Speed	+/-0 kts	A	Ck at roll out
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Major deviation	Angle of bank, speed and/or altitude beyond D	E	

EVALUATOR ASSESSMENT

______

BELOW AVERAGE				AVERAGE			ABOVE AVERAGE				0 ERRORS		
	50		60	70		75	80		85	90	95		100
40		55	65	68	72	76	78	84		88	94	98	

<u>NEXT MANEUVER</u>: MAKE A 180 DEGREE DESCENDING RIGHT RUN AT 1000 FT PER MIN; ROLL OUT HEADING 170 DEGREES AND 5000 FT; MAINTAIN IAS 250 KTS.

Control Number____________

MANEUVER: DESCENDING TURNS (CATEGORIZATION: ROUTINE)
MAKE A 180 DEGREE DESCENDING RIGHT TURN AT 1000 FEET PER MIN; ROLL OUT HEADING 170 DEGREES AND 5000 FT; MAINTAIN IAS 250 KTS.

FUNCTION	CRITERION	SCORE	REMARKS
Turns	30 deg bank	A	
	30 deg bank +/-5 degs	B	
	30 deg bank +/-10 degs	C	
	30 deg bank +/-15 degs	D	
Rate of descent (1000 ft. per minute)	+/-100 ft	A	
	+/-200 ft	B	
	+/-300 ft	C	
Speed	+/-0 kts	A	End of turn
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Heading (roll out)	+/-0 deg	A	
	+/-5 deg	B	
	+/-10 deg	C	
	+/-15 deg	D	
Altitude	+/-0 ft	A	
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Major deviation	Altitude, heading of speed beyond D	E	

EVALUATOR ASSESSMENT ______

BELOW AVERAGE			AVERAGE			ABOVE AVERAGE				0 ERRORS	
	50	60	70		75	80	85	90	95	100	
40	55	65	68	72	76	78	84	88	94	98	

NEXT MANEUVER: ILS RY35R APCH

CLEARANCE: N727 CLEARED DIRECT GALLY; DESCEND AND MAINTAIN 4000; AT GALLY CLEARED FOR ILS RY35R APPROACH AND LANDING AT WILL ROGERS AIRPORT; REPORT GALLY INBOUND.

ATIS: OKC WX E7 BKN 50 OVC 2 70/56 300012 992

Control Number____________

MANEUVER: APPROACH PRECISION ILS35R (CATEGORIZATION: CHALLENGING)
CLEARANCE: N727 CLEARED DIRECT GALLY; DESCEND AND MAINTAIN 4000 FT AT GALLY CLRD FOR ILS RY35R APCH AND LANDING AT WILL ROGERS AIRPORT; REPORT GALLY INBOUND.

OKC WX E7 BKN 50 BKN 2 70/56 300012 992.

FUNCTION	CRITERION	SCORE	REMARKS
Tracking ( ) Radial	0 deviation	A	
(2 1/2 mins)	+/-1 dot	B	
or procedure	+/-2 dots	C	
	+/-3 dots	D	
In range check	Completed	A	
	Omitted	B	
CRM	Exchange info w/ crew; assign duties	A	
	No exchange-delegates	B	
	No crew involvement	C	
Altitude	+/-0 ft	A	
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Turns	30 deg bank	A	
	30 deg bank +/-5 degs	B	
	30 deg bank +/-10 degs	C	
	30 deg bank +/-15 degs	D	
Speed	+/-0 kts	A	
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Procedure turn	Within 10 miles correct procedure	A	
	Beyond pattern	C	
Altitude	+/-0 ft	A	
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Turns	30 deg bank	A	
	30 deg bank +/-5degs	B	
	30 deg bank +/-10 degs	C	
	30 deg bank +/-15 degs	D	

Control Number___________

MANEUVER: APPROACH PRECISION ILS35R (Cont'd)

FUNCTION	CRITERION	SCORE	REMARKS
Speed	+/-0 kts	A	
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Localizer Tracking	0 deviation	A	GALLY
	+/-1 dot	B	
	+/-2 dots	C	
	+/-3 dots	D	
Indicated altitude	+/-0 ft	A	
	+/-50 ft	B	
	+/-100 ft	C	
	+100 ft	D	
Speed	+/-0 kts	A	
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Localizer tracking	0 deviation	A	DH
	+/-1 dot	B	
	+/-2 dots	C	
	+/-3 dots	D	
Glide Slope	Bar centered	A	
	+/-1 dot	B	
	+/-2 dots	C	
	+/-3 dots	D	
Speed	+/-0 kts	A	DH
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Major Deviation	+500 ft alt no localizer stall, etc.	F	

EVALUATOR ASSESSMENT ______

BELOW AVERAGE AVERAGE ABOVE AVERAGE 0 ERRORS

50 60 70 75 80 85 90 95 100

40 55 65 68 72 76 78 84 88 94 98

NEXT MANEUVER: LANDING

Revised page 7/20/92

Control Number__________

MANEUVER: LANDING (CATEGORIZATION: ROUTINE)

FUNCTION	CRITERION	SCORE	REMARKS
Landing checklist	Completed	A	
	Omitted	B	
	Incomplete	C	
Speed	+/-0 kts	A	
	+/-5 kts	B	
	+10 kts	C	
	-10 kts	D	
Runway alignment	Centerline	A	
	1/4 L/R centerline	B	
	+1/4 L/R centerline	C	
Touchdown (1000 ft down runway)	+/-0 ft	A	
	+/-500 ft	B	
	+500 ft	C	
Direction control	Straight	A	Computer
	+/-5 deg	B	
	+5 deg	C	
Reverse thrust	3 of 3 steps	A	
	2 of 3 steps	B	
	1 of 3 steps	C	
After landing checklist	Completed	A	
	Omitted	B	
Major deviation	Stall, off runway, etc.	G	

EVALUATOR ASSESSMENT

______

BELOW AVERAGE AVERAGE ABOVE AVERAGE 0 ERRORS

50 60 70 75 80 85 90 95 100

40 55 65 68 72 76 78 84 88 94 98

Control Number________

EVALUATOR ASSESSMENT OF SCENARIO I

BELOW AVERAGE				AVERAGE				ABOVE AVERAGE				0 ERRORS		
	50		60		70		75		80	85	90	95		100
40		55	65	68		72	76	78		84	88	94	98	

COMMENTS

DEVIATIONS

A ____
B ____
C ____
D ____
E ____
F ____
G ____

TECHNICAL SCORE ____

EVALUATOR ASSESSMENT ____

TOTAL ____

AVERAGE ____

FIRST OFFICER ASSESSMENT ____

Control Number______________

FIRST OFFICER ASSESSMENT OF SCENARIO I

______

BELOW AVERAGE						AVERAGE				ABOVE AVERAGE					0 ERRORS		
	50		60			70		75			80		85		90	95	100
40		55		65	68		72		76	78		84		88	94	98	

COMMENTS

____

____

____

____

____

____

____

____

____

____

____

____

____

APPENDIX B

Results of Questionnaires Used to Assess Subjects' Perceptions of the Simulator Measure and Predictor Tests*

*Descriptive statistics (mean, standard deviation and range) for each item have been noted on each questionnaire.

COGSCREEN QUESTIONNAIRE

The tests in this computerized test battery were not intended to simulate the activities performed while flying. However, they were intended to measure the cognitive functions that underlie safe flying performance. Please read each of the statements below and decide how much you agree or disagree with each by circling the appropriate number. Please be honest and forthright with your responses concerning COGSCREEN.

1. These tests tap into the cognitive abilities that one would expect to find in a safe pilot.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 3.4 SD: 1.1 Range: 2-6

2. This test is too easy and should be made more difficult.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 5.1 SD: 1.3 Range: 2-7

3. This test is a reasonably valid method for screening unsafe pilots.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 4.7 SD: 1.4 Range: 2-7

4. I would not object to taking this test on a routine basis.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 3.6 SD: 2.0 Range: 1-7

5. Do you have any other comments on this test?

6. Please circle the number that best describes your use of personal computers:

1. Novice User 2. Very Inexperienced User 3. Casual User

4. Experienced User 5. Very Experienced User/Programmer

Mean: 2.5 SD: 1.2 Range: 1-5

FLITESCRIPT QUESTIONNAIRE

Flitescript is intended to measure a pilot's memory for ATC communication. There may be a relationship between how much experience a pilot has and how well he performs on the this test. Please read each of the statements below and decide how much you agree or disagree with each by circling the appropriate number. Please be honest and forthright with your responses concerning Flitescript.

1. This test is a good measure of the experience level of a pilot.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 4.7 SD: 2.1 Range: 1-7

2. This test is too easy and should be made more difficult.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 5.6 SD: 1.4 Range: 3-7

3. This test is a reasonably valid method for screening unsafe pilots.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 5.8 SD: 1.4 Range: 3-7

4. I would not object to taking this test on a routine basis.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 4.6 SD: 2.2 Range: 1-7

5. Do you have any other comments on this test?

6. Please circle the number that best describes your experience with the use of a *mouse* in conjunction with a personal computer.

1. Novice User 2. Very Inexperienced User 3. Casual User

4. Experienced User 5. Very Experienced User

Mean: 2.3 SD: 1.4 Range: 1-5

WOMBAT QUESTIONNAIRE

WOMBAT is not intended to measure direct flying skills. It does not depend on flying experience. However, it is intended to measure the activities that underlie flying performance, such as the ability to simultaneously perform several tasks and to deal with constantly changing priorities. Please read each of the statements below and decide how much you agree or disagree with each by circling the appropriate number. Please be honest and forthright with your responses concerning WOMBAT.

1. This test taps into the cognitive abilities that one would expect to find in a safe pilot.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 3.3 SD: 1.3 Range: 1-7

2. This test is too easy and should be made more difficult.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 5.3 SD: 1.4 Range: 2-7

3. This test is a reasonably valid method for screening unsafe pilots.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 4.8 SD: 1.2 Range: 2-7

4. I would not object to taking this test on a routine basis.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 3.7 SD: 1.8 Range: 1-7

5. Do you have any other comments on this test?

6. Please circle the number that best describes your experience with fast-paced video games that use joysticks for control.

1. Novice User 2. Very Inexperienced User 3. Casual User

4. Experienced User 5. Very Experienced User

Mean: 1.8 SD: 1.1 Range: 0-5

SIMULATOR POST-FLIGHT QUESTIONNAIRE

The simulator evaluation is intended to measure pilot proficiency in a realistic flight environment, in both low and high workload situations that involve routine as well as emergency maneuvers. Please read each of the statements below and decide how much you agree or disagree with each by circling the appropriate number. Please be honest and forthright with your responses concerning the simulator evaluation.

1. Performance on this test is a good measure of flying ability.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 1.9 SD: 1.1 Range: 1-6

2. The scenarios are sufficiently comprehensive to make reasonable judgments about overall flying ability.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 1.9 SD: 1.3 Range: 1-6

3. This test is a reasonably valid method for screening unsafe pilots.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 2.5 SD: 1.6 Range: 1-7

4. I would not object to taking this test on a routine basis.

1	2	3	4	5	6	7
Strongly Agree			Not Sure No Opinion			Strongly Disagree

Mean: 2.1 SD: 1.6 Range: 1-7

5. Do you have any other comments on this test?

6. Have you discussed this test with anyone who was a prior test subject? (Yes -- No, No=0). If yes, how much information about scenarios did you know?

1. Nothing 2. Very little 3. Moderate amount
4. Quite a bit 5. Almost everything

Mean: 0.3 SD: 0.6 Range: 0-2

APPENDIX C

Descriptive Statistics

Descriptive Statistics For the Total Sample of 40 Pilots

VARIABLE	MEAN	SD	RANGE
Pilot Age	53.9	8.1	41.0 - 71.0
Total Hours (all types)	14,694.0	7,197.6	5,000.0 - 35,000.0
Total B727 Hours	5,129.8	4,685.1	64.0 - 18,000.0
B727 Hours Last Six Months	156.3	152.4	0.0 - 500.0
B727 Hours Last 30 Days	28.6	41.0	0.0 - 225.0
B727 Simulator Hours	4.4	8.3	0.0 - 32.0
All Simulator Maneuvers:			
Total Deviations	122.4	50.6	34.0 - 356.0
Subjective Evaluation	85.1	4.8	74.7 - 92.5
Routine Maneuvers:			
Total Deviations	43.7	17.7	12.0 - 105.0
Subjective Evaluation	85.0	6.9	58.8 - 93.9
Challenging Maneuvers:			
Total Deviations	39.6	17.8	9.0 - 129.0
Subjective Evaluation	82.4	5.7	69.0 - 92.2
Emergency/Abnormal Maneuvers:			
Total Deviations	39.2	17.7	13.0 - 122.0
Subjective Evaluation	83.6	8.9	44.0 - 93.6
Flitescript:			
% Correct	59.0	14.3	26.7 - 86.7
Wombat:			
Tracking	84.9	6.9	67.3 - 93.9
Bonus	18.9	10.0	0.7 - 38.0
Total	188.9	20.0	142.1 - 223.9
Cogscreen:			
Accuracy	.0	3.4	-9.8 - 6.4
Speed	.1	5.4	-7.7 - 15.7
Total	.0	8.9	-33.1 - 17.6

Descriptive Statistics For the Subgroup of 23 Experienced B727 Pilots

VARIABLE	MEAN	SD	RANGE
Pilot Age	53.0	6.9	41.0 - 71.0
Total Hours (all types)	14,708.0	5,943.9	5,200.0 - 30,800.0
Total B727 Hours	6,409.9	5,264.5	500.0 - 18,000.0
B727 Hours Last Six Months	236.2	134.5	25.0 - 500.0
B727 Hours Last 30 Days	43.2	45.3	0.0 - 225.0
B727 Simulator Hours	7.4	10.6	0.0 - 32.0
All Simulator Maneuvers:			
Total Deviations	104.4	31.2	34.0 - 145.0
Subjective Evaluation	85.8	4.5	77.7 - 92.5
Routine Maneuvers:			
Total Deviations	37.0	12.6	12.0 - 55.0
Subjective Evaluation	87.1	4.3	79.4 - 93.9
Challenging Maneuvers:			
Total Deviations	34.8	10.7	9.0 - 57.0
Subjective Evaluation	83.5	5.1	73.8 - 92.2
Emergency/Abnormal Maneuvers:			
Total Deviations	33.4	11.5	13.0 - 51.0
Subjective Evaluation	86.0	5.2	75.2 - 93.6
Flitescript:			
% Correct	58.6	14.6	26.7 - 80.0
Wombat:			
Tracking	84.7	6.4	67.3 - 93.9
Bonus	18.4	11.0	0.7 - 38.0
Total	188.1	21.1	142.1 - 223.9
Cogscreen:			
Accuracy	0.0	3.9	-9.8 - 6.4
Speed	0.3	5.7	7.7 - 15.7
Total	-0.2	10.3	-33.1 - 17.6

★U.S. GOVERNMENT PRINTING OFFICE: 1994 - 660-067/00062