NAVAL POSTGRADUATE SCHOOL
Monterey, California

19980909 016

THESIS

DETERMINANTS OF FLIGHT TRAINING PERFORMANCE: NAVAL ACADEMY CLASSES OF 1995 AND 1996

by

Paul M. Reinhart

August 1998

Thesis Co-Advisors: Gregory G. Hildebrandt
Roger D. Little

Approved for public release; distribution is unlimited.

DTIC QUALITY INSPECTED 1

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE August 1998	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE

DETERMINANTS OF FLIGHT TRAINING PERFORMANCE: NAVAL ACADEMY CLASSES OF 1995 AND 1996.

5. FUNDING NUMBERS

6. AUTHOR(S)
Reinhart, Paul M.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

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

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. **ABSTRACT** This thesis investigates the relationship between observable characteristics and performance during the primary phase of flight training. The data for this study consists of 272 observations from Naval Academy graduates in the classes of 1995 and 1996. Analysis of the variables was conducted using the Heckman two-stage regression technique to correct for possible selectivity bias. In this technique the first-stage probit model, which predicts the likelihood of primary phase completion, is used to generate a correction factor for possible selectivity. The correction factor is then used in the second-stage adjusted least-squares regression model. The conclusions from this study are: The biographical inventory from the Aviation Selection Test Battery(ASTB) is a valid predictor of primary phase completion. The Pilot Flight Aptitude Rating(PFAR) from the ASTB, academic achievement(AQPR) at the Naval Academy, and previous flight experience are all valid predictors of flight training performance. Additionally, it appears that sample selection bias does not seem to be a problem in this analysis.

14. SUBJECT TERMS
Naval Academy, Pilot Selection, Flight Training, Selectivity Bias

15. NUMBER OF PAGES
96

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFI- CATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited.

DETERMINANTS OF FLIGHT TRAINING PERFORMANCE: NAVAL ACADEMY CLASSES OF 1995 AND 1996

Paul M. Reinhart
Lieutenant, United States Navy
B.S., United States Naval Academy, 1991

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE
IN
LEADERSHIP AND HUMAN RESOURCE DEVELOPMENT

from the

NAVAL POSTGRADUATE SCHOOL
August 1998

Author: _____
Paul M. Reinhart

Approved by: _____
Gregory G. Hildebrandt, Thesis Co-Advisor

_____
Roger D. Little, Thesis Co-Advisor

_____
Reuben T. Harris, Chairman
Department of Systems Management

ABSTRACT

This thesis investigates the relationship between observable characteristics and performance during the primary phase of flight training. The data for this study consists of 272 observations from Naval Academy graduates in the classes of 1995 and 1996. Analysis of the variables was conducted using the Heckman two-stage regression technique to correct for possible selectivity bias. In this technique the first-stage probit model, which predicts the likelihood of primary phase completion, is used to generate a correction factor for possible selectivity bias. The correction factor is then used in the second-stage adjusted least-squares regression model. The conclusions from this study are: The biographical inventory from the Aviation Selection Test Battery(ASTB) is a valid predictor of primary phase completion. The Pilot Flight Aptitude Rating(PFAR) from the ASTB, academic achievement(AQPR) at the Naval Academy, and previous flight experience are valid predictors of flight training performance. Additionally, it appears that sample selection bias does not seem to be a problem in this analysis.

TABLE OF CONTENTS

I. INTRODUCTION 1

A. INVESTMENTS IN CAPITAL 2

B. SCOPE 5

C. HYPOTHESES 6

D. ORGANIZATION 7

II. PILOT SELECTION METHODS 9

A. NAVY 9

1. Aviation Selection Test Battery 9

2. Naval Academy 11

B. AIR FORCE 13

1. Air Force Officer Qualifying Test 14

2. Air Force Academy 15

3. Air Force Reserve Officer Training Corps 16

4. Officer Training School 17

C. CANADIAN AIR FORCE 18

D. LUFTWAFFE 19

E. LUFTHANSA 22

F. AMERICAN CIVILIAN CARRIERS 23

III. HISTORY OF PILOT SELECTION RESEARCH 25

A. GENERAL COGNITIVE MEASURES 26

B. PSYCHOMOTOR MEASURES 31

C. PERSONALITY MEASURES 34

D. JOB SAMPLE MEASURES 35

E. INTEGRATING THE RESEARCH FINDINGS 37

IV. CONSTRUCTION OF PROBABILITY MODEL 41

A. DATA SOURCES 41

B. ANALYSIS OF VARIABLES 42

1. Dependant Variables 42

2. Independent Variables 45

C. MODEL DESCRIPTION 50

1. Selectivity Bias 51

2. Censored Regression Model 52

D. TWO-STAGE HECKMAN PROCEDURE 55

V. ANALYSIS OF EMPIRICAL MODELS 57

A. FIRST-STAGE PROBIT MODEL 58

1. Goodness of Fit 62

2. Notional Pilot 62

B. SECOND-STAGE OLS MODEL 64

VI. CONCLUSIONS AND RECOMMENDATIONS 73

A. CONCLUSIONS 73

B.. RECOMMENDATIONS FOR FUTURE RESEARCH 77

REFERENCES 81

INITIAL DISTRIBUTION LIST 85

I. INTRODUCTION

The selection of naval aviators is serious business. The training program is a long, arduous, and expensive process, often requiring in excess of two years to complete. The inherent difficulty of the program is necessitated by the demands that will eventually be placed upon these individuals. Navy and Marine Corps pilots are required to operate under conditions that are largely unique to the sea-going services such as shipboard operations where the nearest alternate landing sight could be several thousand miles away. Despite these demands, naval aviators continue to operate around the world conducting day-and-night flight operations in conditions that would cause most prudent civilian aviators to remain safely on the ground.

The sudden collapse of the Soviet Union in the early 1990's brought to an end the high levels of military spending that had been sustained during the Cold War era. Lacking a specific foe to concentrate on, the perceived threat from external forces diminished and the American public called for a "peace dividend." In response to public opinion, Congress continues to support reduced levels of military spending. The military budget for 1997 was set at around 265 billion dollars, and as we approach the turn of the century, will remain at this level, with only marginal

increases and zero real growth in spending for the foreseeable future. Reduced budgets have not necessarily resulted in reduced commitments. Military leaders are still required to plan, equip, and train forces to fight two simultaneous Major Theater wars. The implications for military leaders, given these budgetary constraints, are that they need to ensure maximum return on every dollar spent.

A. INVESTMENTS IN CAPITAL

Navy leaders must seek to optimize investments in both physical and personnel capital in order to maximize military effectiveness. Investments in military physical capital include ships, aircraft, and bases. Effectiveness can be maximized by examining which is the most appropriate weapons system, which system has the lowest initial acquisition cost, and which system has the lowest life cycle operating and support cost. Investments in military physical capital are highly political and are often driven by forces beyond the control of Navy leaders. Investments in military personnel capital take the form of educating or training personnel to enhance performance. Examples of investments in human capital include post-graduate education and Navy training schools such as nuclear power school and flight training.

Flight training is one of the most expensive training opportunities provided by the Navy. The training costs per individual range from $350,000 for a helicopter pilot to $700,000 for a jet pilot (Bowman, 1995). Bowman points out that these estimates do not take into consideration the costs of the military flight instructors, so the actual costs are much higher. An additional training cost associated with these individuals is the cost to make selected pilots proficient in a particular warfare specialty. This additional training requires a substantial investment of both resources and time.

The process of pilot selection can help maximize the Navy's investment in personnel capital by selecting those individuals with the highest probability of completing flight training, thereby reducing the high costs associated with attrition from the flight program. The problem of attrition from flight training is neither a new phenomenon, nor is it unique to the Navy. It exists for every military organization that conducts flight training and is a major concern because the attrition cost per individual increases as the time spent in the program increases. The cost has recently been estimated at $18,000 for an individual who attrites from Aviation Pre-flight Indoctrination (API), $50,000 while in Primary, $200,000 for Advanced Helicopter,

and up to $500,000 for individuals in Advanced Jet training(Blower, 1997). Attrition has the additional side effect of hindering the Training Command's ability to meet the manning requirements of fleet squadrons. If attrition is higher than expected, a serious shortfall in the operational manning level of the fleet squadrons could develop, resulting in reduced effectiveness and a weakening of national defense.

In addition to the direct cost of attrition, the opportunity cost of training opportunities lost to those who attrite must be accounted for when assessing the total cost involved in a flight failure. Consider the following hypothetical example. A Midshipman selected for aviation training graduates from the Naval Academy and is given a six month temporary assignment while waiting to report to Pensacola to begin flight training. The individual reports to Pensacola, completes Aviation Indoctrination and begins flight training.

Unfortunately, this student attrites in the 14th week of primary training. This individual is redesignated and sent to another community(e.g., Surface Line or Supply Corps) to begin training anew. If this flight failure had been predicted, the student in question could have been sent directly to a different community for training and would

have reported to the fleet to relieve another officer almost a year earlier.

As can be seen by looking at the bigger picture, the costs associated with a flight failure are actually much higher than simply the resources expended on a particular individual while in the flight program. The research question to be addressed in this thesis is: Are there identifiable and significant characteristics that can be used to improve the pilot selection process in use at the Naval Academy?

B. SCOPE

The scope of this research project is limited to Naval Academy graduates from the classes of 1995 and 1996. To date, 1995 and 1996 represent the only classes that have taken the latest version of the Aviation Selection Test Battery(ASTB) for which flight training performance data are available and have been subject to the new service assignment policy. Graduates from these classes who were selected for student naval flight officer training were not included due to differences in the training curriculum following Aviation Indoctrination. There was no attempt to include peer student pilots from other accession sources, who had also taken the latest version of the ASTB, due to the extreme difficulty in collecting relevant data.

Data for this thesis were compiled from a variety of sources including the Naval Operational Medicine Institute(NOMI), Pensacola Florida, Naval Personnel Records Data Center, San Diego, and Institutional Research, United States Naval Academy. Data sets were combined using Excel spreadsheets and all statistical analysis was conducted using SPSS 7.5.

C. HYPOTHESES

This research paper will test the following hypotheses:

- Naval Academy graduates who score higher on the biographical inventory are more likely to complete primary flight training than those with lower scores.
- Naval Academy graduates who score higher on the pilot flight aptitude rating achieve higher flight grades that those with lower scores.
- Naval Academy graduates with higher academic quality point ratings achieve higher flight grades than those with lower ratings.
- Naval Academy graduates with higher military quality point ratings achieve higher flight grades than those with lower ratings.
- Naval Academy graduates with technical degrees achieve higher scores that graduates with non-technical degrees.
- Male graduates achieve higher flight grades than female graduates.
- Ethnic majority graduates achieve higher flight grades than minority graduates.
- Naval Academy graduates with previous flight experience achieve higher flight grades than graduates with no previous flight experience.

D. ORGANIZATION

Following the introduction, Chapter II will look at various pilot selection methods used by the Navy, Air Force and a selection of other organizations as a means of acquainting the reader with the numerous and varied methods utilized in selecting pilot candidates. Chapter III will present a history of pilot selection. Chapter IV will discuss the methodology and models to be used in analyzing the data. Chapter V will present and interpret the results of the models. Finally, Chapter VI will summarize the conclusions of this study and offer recommendations for future research.

II. PILOT SELECTION METHODS

A. NAVY

The Navy has several sources from which to draw potential student naval aviators; the primary three are Officer Candidate School(OCS), Reserve Officer Training Corps(ROTC), and the Naval Academy(USNA). Regardless of accession source all student naval aviators must meet certain basic requirements which include:

- Hold an accredited bachelor's degree.
- Be at least 19 but less than 27 years of age at the time of commissioning.
- Be physically qualified and aeronautically adaptable according to Navy standards, with 20/20 uncorrected vision and normal color and depth perception.
- Meet anthropometric standards which determine if the aviator can safely fit in an aircraft cockpit.
- Achieve a qualifying score on the Aviation Selection Test Battery(ASTB).

1. Aviation Selection Test Battery

The 1992 Navy and Marine Corps Aviation Selection Test Battery(ASTB) is the most recent revision in a series of cognitive pencil-and-paper tests that have been used by the naval services as a selection measure for student naval aviators. This test battery replaced the Academic

Qualifying Test/Flight Aptitude Rating(AQT/FAR) that was first instituted in 1942 with revisions in 1953 and 1971. Reasons for the latest revision include: changes in demographics due to the all-volunteer-force, changes in naval aviation training, decrease in predictive validity of the previous test, possible compromise of the test since its last revision, and changes in federal guidelines regarding employee selection procedures.

Developed by Educational Testing Services of Princeton, New Jersey, in conjunction with the Naval Aerospace and Operational Medical Institute, the ASTB is composed of six subtests: Math-Verbal Test(MVT), Mechanical comprehension Test(MCT), Spatial Apperception Test(SAT), Aviation and Nautical Test(AN), Biographical Inventory(BI), and Aviation Interest(AI). Weighted combinations of the various subtests result in five stanine scores(a standard score on a nine point scale), that are used to predict attrition, academic performance, and basic flight performance. These scores are: Academic Qualifications Rating(AQR), Pilot Flight Aptitude Rating(PFAR), Flight Officer Aptitude Rating(FOFAR), Pilot Biographical Inventory (PBI), and Flight Officer Biographical Inventory(FOBI).

The ASTB differs from previous tests in that only the PBI and FOBI are intended to predict attrition. The AQR

predicts academic performance, and the PFAR and FOFAR predict basic flight performance(Frank & Baisden, 1993). Navy pilot applicants must obtain qualifying scores of 3/4/4 on the AQR, PFAR, and BI. Marine pilots must achieve scores of 4/6/4. Once these basic requirements have been met, each accession source uses a unique process to fill its allotment of primary flight training billets.

2. Naval Academy

For over 150 years the Naval Academy has been a primary source for career officers. The percentage of naval officers who are academy graduates has fallen dramatically since WWII and the inception of the ROTC and OCS programs, but the academy remains the largest single commissioning source accounting for 30% of the pilot community(Bowman, 1995). The historical overall attrition rate from Navy flight training has been near 30% but Naval Academy graduates have faired slightly better than student naval aviators from the other two accession sources(North & Griffin, 1977). This has been attributed to the fact that the selective nature of the admissions process combined with a rigorous four year course of study acts as a effective filter to screen for adaptability to a military environment (Griffin & Mosko, 1977).

Admission to the Naval Academy is a long and selective process usually involving a congressional nomination and what is known as the whole-person philosophy which generates a candidate multiple based on academic, athletic, and civic accomplishments. While at the Naval Academy, Midshipmen undergo a rigorous four year academic and military curriculum. The culmination of these four years is service assignment. Prior to 1995 Midshipmen selected their desired community based solely on Order of Merit and physical qualification. The Order of Merit was determined by academic, military performance, and conduct grades, with academic grades the primary determinant.

The new selection process involves an interview phase, community screening phase, preference designation phase, and an assignment phase that is designed to make the process even more selective. During the interview phase each graduating Midshipman is interviewed by a team consisting of two or three commissioned officers from the different service communities. The interview is then scored on a scale of zero to ten, with zero being unsatisfactory and ten representing excellent, in five areas including appearance and poise, oral communication, leadership potential, service community motivation, and service community understanding.

The community screening phase is designed to assess

medical qualification and community specific academic and physical requirements. For potential student naval aviators this requirement includes the minimum score on the Aviation Selection Test Battery(ASTB) described earlier. Preference designation phase is the process during which each First-class Midshipman will officially indicate their service and community preferences from among those which they have been found qualified. During the final assignment phase, service assignment boards from each community will select the best qualified Midshipmen from among those applying.

B. AIR FORCE

The Air Force operates a pilot training program similar to that of the Navy. They also rely on three primary accession sources, the Air Force Academy, AFROTC, and Officer Training School, to provide student pilots for Specialized Undergraduate Pilot Training(SUPT). The requirements for SUPT are that the candidate be less than 27 and ½ years old, have a bachelor's degree, pass a physical examination, obtain satisfactory scores on the Air Force Officer Qualifying Test(AFOQT) and Basic Attributes Test(BAT), and successfully complete a flight screening program.

1. Air Force Officer Qualifying Test

The Air Force Officer Qualifying Test(AFOQT) is a paper-and-pencil multiple aptitude battery similar to the Navy ASTB. This test is used for officer commissioning and aircrew training selection. The test has been in use since 1957, with new forms being developed about every seven years. The current version of the AFOQT is Form O.

The current AFOQT battery has five composite scores: Verbal, Quantitative, Academic Aptitude, Pilot, and Navigator-Technical. Its overall purpose is to test general cognitive ability,(g), and five lower order functions: verbal, math, spatial, interest/aptitude, and perceptual speed. The 16 tests included in battery are: Verbal Analogies(VA), Arithmetic Reasoning(AR), Reading Comprehension(RC), Data Interpretation(DI), Word Knowledge(WK), Math Knowledge(MK), Mechanical Comprehension(MC), Electrical Maze(EM), Scale Reading(SR), Instrument Comprehension(IC), Block Counting(BC), Table Reading(TR), Aviation Information(AI), Rotated Blocks(RB), General Science(GS), and Hidden Figures(HF)(Carretta, 1997).

The Basic Attributes Test(BAT) is a computer-based test battery used to augment the results of the AFOQT for pilot selection. It consists of five tests that measure psychomotor coordination, short-term memory, and attitudes

towards risk taking. The BAT tests are the Two-hand Coordination, Complex Coordination, Item Recognition, Time Sharing, and Activities Interest Inventory.

2. Air Force Academy

The Air Force Academy located in Colorado Springs, Colorado, is in many ways comparable to the Naval Academy. It is a four year institution that provides students with total immersion in a military culture and has a similarly competitive admissions process that normally requires a congressional nomination. Approximately 65 percent of each Air Force Academy graduating class will enter Specialized Undergraduate Pilot Training(SUPT). This represents approximately 25 percent of all pilot candidates who enter SUPT each year. During the course of instruction at the Academy, cadets participate in the Pilot Indoctrination Program(PIP).

The stated purpose of this program is "to identify students who possess the potential to complete SUPT and motivate qualified Academy graduates toward a rated career in the Air Force." This program consists of seven hours of airmanship academics and 20 hours of light single engine training. PIP also has an optional phase which consists of 4.5 hours of incentive sorties. Upon graduation each candidate who has successfully completed PIP, is medically

qualified, receives a positive recommendation from a flying supervisor, and has a desire to attend pilot training will receive a SUPT assignment. Air Force Academy graduates are exempted from taking the Air Force Officer Qualifying Test(AFOQT).

3. Air Force Reserve Officer Training Corps

Air Force Reserve Officer Training Corps(AFROTC) provides the largest number of aviation candidates, accounting for over one-third of SUPT entrants. ROTC candidates come from over 150 colleges and universities from across the United States. In order to compete for an ROTC SUPT assignment, candidates must pass a medical evaluation and achieve a minimum score on the AFOQT. The minimum percentile score on the pilot composite is a 25 and 10 on the navigator-technical composite. In addition, the combined score must total at least 50 (Lynch, 1991). Other factors taken into consideration are grade point average, Scholastic Aptitude Test scores, Basic Attributes Test(BAT) score, and unit commander ratings. Selection for ROTC SUPT quotas is determined by a central selection board that convenes semi-annually at Maxwell AFB, Alabama.

Candidates approved by the selection board who do not already possess a private pilots license are required to complete Light Aircraft Training(LATR) between their junior

and senior year at the USAF Officer Training School flight screening facility at Hondo AFB, Texas. LATR includes 13 hours of academic instruction and 14 hours of light aircraft training. Successful completion of the light aircraft training marks the end of the flight screening phase and candidates enter SUPT following graduation and commissioning.

4. Officer Training School

Air Force Officer Training School(OTS) is a 120 day course conducted at Lackland AFB, Texas, that provides both officer training and flight screening. OTS production rates are more flexible than the Air Force Academy or the ROTC program and are therefore used as a buffer to accommodate changing SUPT production rates. Acceptance as a cadet requires a college degree, passing a medical examination, and obtaining a qualifying scores on the AFOQT and BAT. Officer Training School quotas are filled by the OTS selection board that is convened "as required," currently six times per year, at Randolph AFB, Texas.

This board applies what is known as the Pilot Candidate Selection Method(PCSM) to rate each applicant. The PCSM uses the results of the AFOQT, BAT, college GPA, and previous flight hours to generate a whole person score for each applicant. The highest scoring applicants, up to the

quotas available, will be accepted for OTS. Those not selected will return to the pool to await the next selection board.

OTS candidates who do not already possess a private pilot license are sent to Hondo AFB, Texas to complete the OTS Flight Screening Program(FSP). Training is 16 days in length and includes 13 hours of academic training and 14 hours of light aircraft training. The culmination of the course is a solo flight and a final evaluation flight. Candidates are scored as either satisfactory or unsatisfactory. Those candidates receiving a satisfactory score begin OTS and upon completion will attend SUPT.

As can be seen from the descriptions of Navy and Air Force selection methods, they rely heavily on the use of paper-and-pencil general cognitive measures to select pilot candidates. The major difference between the services is that the Air Force places an additional emphasis on light plane screening while the Navy does not. In the remaining sections of this chapter, I will briefly look at other types of selection methods in use.

C. CANADIAN AIR FORCE

The pilot selection process by which the Canadian Armed Forces assesses pilot applicants is known as the Canadian Automated Pilot Selection System(CAPSS). CAPSS is a stand-

alone selection devise that provides a measure of complex cognitive abilities and psychomotor coordination. The underlying constructs that CAPSS is designed to measure are: psychomotor coordination, learning rate, multi-task integration, and performance under-overload(Adams-Roy, 1996).

CAPSS uses flight simulator technology to collect data on a sample of flight skills required to pilot a light single engine aircraft. The syllabus consists of five one-hour sessions where each segment is an extension of, and uses skills learned in, the previous session. At the end of each segment a score, ranging from 0 to 1 is generated. This score represents the probability of a candidate's successful completion of flight training. The best correlation with success in flight training has come from the score after the end of the fourth segment, and it is this score which is used to select pilot candidates.

D. LUFTWAFFE

The German Air Force, or Luftwaffe, uses a multi-stage psychological selection process to select from potential aviation candidates. This process, administered by Division IV of the Luftwaffe Institute of Aerospace Medicine(FMI), can last from between one and two years. The division consists of four branches: Selection, Screening,

Diagnostics, and Experimental Psychology. The psychological selection process is organized into three sequential steps: Pre-selection, Main Selection, and Flight Screening. Every applicant must successfully complete all steps to be admitted into pilot training(Gnan, Flyer, and King, 1995).

During the Pre-Selection stage any male who wishes to become a Lufftwaffe pilot may apply. The average age of applicants to the aviation service is about 19. Applicants will first take the Officer Qualification Test which includes objective testing, interviews and a physical stamina test. Those who pass the test will be sorted into three groups through the use of an intelligence test and a concentration test. Applicants from the top two groups are accepted for the next selection step. The lowest performing group may be accepted in cases where there are an insufficient number of applicants in the higher performing groups.

The Main Selection stage is a one-day examination conducted by the Selection and Diagnostic Branches. This examination consists of three computer-based psychomotor test, a biographical inventory, and a diagnostic interview. The psychomotor tests measure coordination, selective attention, reaction time and accuracy, multiple task performance, stress tolerance, and decisiveness. The

diagnostic interview focuses on stress reaction during the tests, coping strategies, achievement, and flying motivation. When the tests are complete the results are reviewed by a board of psychologists to determine if the applicant will be accepted to the next selection step. Once approved by the selection board, candidates are given a flight physical medical evaluation. During the Main Selection phase almost 30% of the applicants are rejected. Having passed Pre-Selection, Main Selection, and physical screening, applicants will be sent to Primary Military Training and Officer's Training School before beginning the final phase of Flight Screening.

Flight Screening begins on the FPS-80 simulator devise. Applicants will fly several missions while their performance and behavior is evaluated. This is followed by 70 hours of academic training covering navigation, meteorology, aerodynamics, and airmanship. The final step in the screening process is 18 hours of flight time in a light single engine aircraft, culminating in a solo flight. Having successfully completed all stages of the selection process, candidates will be sent to either EURO-NATO Joint Jet-Pilot Training at Sheppard AFB, Texas, or Weapons Systems Officer training at Randolph AFB, Texas, depending on the amount of points accumulated during Flight Screening.

E. LUFTHANSA

The Department of Aviation and Space Psychology of the German Aerospace Research Establishment(DLR) selects pilot applicants for *ab-inito*, no previous flying experience required, training for Lufthansa Airlines. Pilot applicants must have completed the German university entrance level(Abitur), and not be more than 27 years of age. The selection process is completed in two stages that evaluate flying and managerial skills.

The first stage lasts two days and applicants are required to perform group administered performance and personality tests designed to measure: technical knowledge and comprehension, mathematics, concentration, speed perception, spatial orientation, and temperament characteristics. Personality assessment is done through the Temperament Structure Scales(TSS) designed by the DLR specifically for pilot selection. The TSS is a multidimensional personality questionnaire that reports 10 personality dimensions measured by stanine scales: work-related traits(motivation, rigidity, mobility, and vitality), social-behavior traits(extroversion, dominance, and aggressiveness), and stress resistance factors(emotional stability, spoiltness, and empathy)(Gnan, et al., 1995). This phase of screening usually reduces the number of

applicants by 70-75%.

The second phase lasts three days and consists of psychomotor coordination and multiple task capacity tests, an interview in front of a selection board, and a medical examination. Applicants who pass the second phase, usually about 10% of the original group, are sent to two years of pilot training in either Bremen, Germany or Phoenix, Arizona. Upon completion of training the student holds all relevant airline licenses including the Airline Transport Pilot License(ATP). After a five month transition period that includes a type rating and a line check, the new pilot is certified to fly as first officer in Lufthansa's entry fleet of Boeing 737s.

F. AMERICAN CIVILIAN CARRIERS

Lufthansa differs from most large American air carriers in that most of their pilots have no previous flight experience and are trained in-house. For decades American air carriers have relied on hiring surplus military pilots with significant experience and training in either the Navy or Air Force. The luxury of being able to fill their ranks with military trained pilots is rapidly becoming a thing of the past for American carriers. This is a result of a decrease in the number of surplus pilots available, rapid expansion throughout the airline industry, and forced

retirement of a large number of Vietnam era pilots. This imbalance in the supply and demand of surplus military pilots is forcing the American carriers to hire increasing numbers of civilian trained pilots and may eventually force them into hiring pilots for *ab-inito* training.

III. HISTORY OF PILOT SELECTION RESEARCH

Orville and Wilbur Wright employed the initial pilot selection measure to determine who would be the first to test their new invention, the airplane. The selection measure they chose was a simple flip of a coin. Wilbur won the toss, but due to mechanical problems his flight lasted less than four seconds and covered less than 100 feet. The Flyer was repaired, and five days later it was Orville's turn. His flight off that small hill in Kitty Hawk on 17 December, 1903 is generally recognized as the watershed event that ushered in the age of flight(Cope, 1996). Aviation technology has evolved greatly since this historic day nearly a century ago, and so have the needs and methods of pilot selection.

The literature on pilot selection is largely focused on military pilots due to the military requirement to train large numbers of pilots, especially during times of war, and the considerable resources available to conduct research. Burke & Hunter(1990) noted during a recent study on pilot selection that of 254 research studies found, only seven dealt with the selection of pilots for non-military settings. The vast literature on pilot selection can be generalized into four main categories: paper-and-pencil

general cognitive measures, psychomotor measures, personality measures, and job sample measures. These categories are explored in turn below.

A. GENERAL COGNITIVE MEASURES

The field of aviation psychology was born during World War I as an offshoot of the general military recruit testing program. When the United States entered the war, the armed forces had fewer than 500 pilots within their ranks, but by the wars end the number exceeded 19,000(Griffin & North, 1977). During this tremendous buildup there was a requirement to provide a nearly continuous stream of new pilots to meet the demands of the war effort. Staggering attrition rates during training, often as high as 50 to 90 percent in the early years of the war, drew the attention of military leaders and scientists(Hunter & Burke, 1995).

In an effort to reduce the high levels of attrition, scientifically designed tests were constructed to select pilot candidates. These earliest tests, designed by Bachman(1918) and Stratton, McComas, Coover, and Bagby(1920), were very rudimentary by today's standards. Without previous knowledge to draw upon, these scientists designed tests for abilities, such as judgment of distance, speed, and time, they thought would be important to flying.

Bachman designed tests for naval aviators which

involved the estimation of the length of sticks, the relative speed of revolving disks, and the time for sand to flow from one container to another. Stratton, et al., refined these tests by including tasks such as judgment of curves and route memory though the use of a cardboard maze in the selection of Army Air Service pilots. It was determined that these tests were correlated with flight training success.

The United States was the only country to use general intelligence testing as a means of pilot selection during World War I. Hermon(1919) conducted a study using the Thorndike Intelligence Test in the selection of Army Air Service Pilots. This test was given to 150 pilots divided into three groups of 50 with the following classifications: flying ability rated as very good, flying ability rated as very bad, and unknown flying ability. A correlation of .35 was found between the Thorndike Intelligence Test and these rough measures of flying ability.

In a later review of this study, Hunter and Burke(1990) note that the use of extreme groups inflated the observed correlations and the statistical significance could not be calculated. The study is however an early indication of a relationship between general intelligence and flying ability. the military's interest in pilot selection

research waned in the post-war period as many of the psychologists who had been involved in this research returned to private practice.

As World War II loomed, U.S. interest in pilot selection research was renewed. Building on the theme of general intelligence testing, and expanding on it to include mechanical comprehension, tests were designed under the Civilian Pilot Training Program, sponsored by the Civil Aeronautics Administration. These tests along with an aviation classification test were later adopted by both the Army and the Navy in the selection of pilot candidates. The aviation selection test adopted by the Navy eventually became the first version of the Aviation Selection Test Battery(ASTB).

Fiske(1947) evaluated these tests on three separate samples of Navy pilot candidates in order to validate their effectiveness. He reported that mechanical comprehension tests had consistently higher correlation with flight performance than did general intelligence as measured by Wonderlic s Personnel Test. Melton(1947) found similar correlation for mechanical comprehension in a study of U.S. Army Air Corps pilots candidates.

Some 20 years later, Berkshire(1967) tried to identify critical attributes that the selection battery failed to

assess in a study of naval aviation candidates who had scored high on the Navy selection battery, but failed to complete training. His study produced three new tests: altitude judgment, maneuver tests, and instrument comprehension. Subsequent validation tests were conducted for altitude judgment and instrument comprehension. Berkshire failed to find correlation for flight training success and the altitude judgment test but revealed a correlation for the instrument comprehension test. Berkshire found, as Fiske(1947) and Melton(1947) had earlier, that the best single predictor of success in his test battery was the mechanical comprehension test.

In 1971 the Navy revamped the selection test battery. The individual components with weighted scores were combined to form a set of composite scores known as the Academic Qualifying Test (AQT) and the Flight Aptitude Rating(FAR). These composite scores, with minor modifications under the current ASTB, remain the foundation upon which student naval aviators are selected. North and Griffin(1977) conducted an analysis of the 1973 cohort of Navy pilot trainees which found significant correlation for individual components of the FAR in the areas of mechanical comprehension, spatial apperception, and biographical inventory.

Like the Navy, both the Army and Air Force have an

extensive battery of written tests used for pilot selection. The Army has the Flight Aptitude Selection Test(FAST), a compilation of eight subtests on biographical information, instrument comprehension, mechanical principles, complex movements, stick and rudder orientation, helicopter information, and flight planning. And finally, the Air Force has the Air Force Officer Qualifying Test(AFOQT). Olea and Rea(1993) and Carretta and Rea(1994) have closely examined the predictive validity of the most recent version of the AFOQT. These studies have explored the composition of the various subtests in terms of general cognitive ability, (g), and specific ability,(s). As was previously found, g-loaded sections of the AFOQT were the best predictors of success in flight training.

The use of pencil-and-paper tests of cognitive ability in pilot selection is not confined to the United States. Burke(1993) reports that tests of spatial orientation, mathematical and verbal reasoning, instrument comprehension, mechanical principles and general intelligence are widely used. NATO countries as well use some or all of these tests in their pilot selection process. While these test have shown little improvement in prediction validity since they were first introduced during WWII, they remain the backbone of U.S. pilot selection measures because of their relatively

low cost per individual and the ease of administration at decentralized testing centers.

B. PSYCHOMOTOR MEASURES

The use of apparatus-based testing measures as a means of pilot selection has existed since World War I. They have however experienced periods of disfavor, especially in the United States, due to problems associated with calibrating the machinery and decentralized testing sites. Early tests conducted during World War I were rather rudimentary in nature and lacked a significant scientific foundation as a basis for inclusion in the selection process. Henmon(1919), Stratton, et al.(1920), and Dockery and Isaacs(1921) performed tests which measured simple reaction times, but these tests indicated this was not a valid predictor of pilot ability(Burke, 1990).

Studies conducted during and after World War II to evaluate pilot performance saw great improvement in the area of scientific rigor. Melton(1947) conducted a series of tests for the U.S. Air Force which evaluated the validity of the Complex Coordination Test in predicting pilot success. This test, built upon earlier research by Reid(1924) and Mashburn(1934), used an apparatus which was configured to resemble an aircraft and consisted of four groups of electric lamps surrounding a speed indicator. These lamps

would illuminate to indicate the simulated orientation of the aircraft when it departed from straight and level flight. The goal was to manipulate the stick and rudder to return to straight and level flight thereby extinguishing the lamps. Using this test, correlation with a dichotomous pass/fail criteria on completion of training, ranged from 0.19 to 0.45.

Fleishman(1956) conducted similar tests to which he added the Rudder Control Test, Two-handed Coordination Test, Directional Control Test, and Rotary Pursuit Test for which he found significant correlation with completion of flight training. His conclusion was that the Complex Coordination Test was the best of these psychomotor tests as a predictor of pilot training success. Despite the early promise of psychomotor tests, the Navy never utilized these tests in the selection of aviators. It was Navy policy not to use test devices unless they could be administered easily and inexpensively at decentralized test stations(North & Griffin, 1977). The Air Force abandoned its apparatus testing program in the early 1950's because of problems with maintaining and calibrating test equipment.

The dawn of the computer age has given new life to the use of psychomotor testing in pilot selection. The most recent computer-based test developed by the Air Force is the

Basic Attributes Test(BAT). This new device was designed to use microprocessor technology in the administration of the Two-hand Coordination Test and the Complex Coordination Test. The BAT was successfully validated by tests at the Air Force Human Resources Laboratory, and has been expanded to 12 subtests. Kantor and Carretta (1988) conducted a cross-validation of the BAT, that produced a correlation of 0.18 using a dichotomous pass/fail variable. The practical utility of the test was demonstrated by showing that using the 10th percentile as a cut-off score would effectively reduce attrition by 20%. It would also however reject 8% of graduates.

Blower and Dolgin(1991) continued the Navy's effort to identify valid psychomotor tests. An evaluation of a recently developed test battery designed by the Naval Aerospace Medical Research Lab, reports significant correlation results for three of the tests, Absolute Difference-Horizontal Tracking, Complex Visual Information, and Risk Taking. These test were found to be generally equivalent in the prediction of pilot training success and they recommend that any one of the three be used in conjunction with the ASTB to increase the validity of the selection model.

Hunter and Burke(1994) found that psychomotor and other

apparatus-based tests are among the best predictors of pilot training success. As computer technology advances it is likely that more realistic, portable, and cost-effective computer-based psychomotor tests will evolve.

C. PERSONALITY MEASURES

The use of personality measures as a means of pilot selection has met with only limited success. Early attempts suffered from the same conditions that afflicted early psychomotor studies, namely lack of scientific rigor. Rippon and Manual(1918) described the successful pilot as a "sportsman who seldom takes his work seriously but looks upon Hun-strafing as a great game." Dockery and Isaacs(1921) thought "quiet, methodical men are among the best fliers." These assessments reflected the general belief that underlying personality characteristics were in some way associated with becoming a successful pilot(Burke, 1990).

Some success with the use of personality measures came during World War II with the development of biographical inventories, the use of which continues today as part of the AFOQT and the Navy ASTB. In a review of Navy selection research, Griffin and Mosko(1977) found almost 40 different personality measures that had been evaluated and failed to make significant contributions to the selection process.

They attributed much of the failure to test bias. All the studies reviewed involved the selection of student naval aviators, a group they contend is highly intelligent and susceptible to response faking. Dolgin and Gibb(1989) also comment on the failure of personality measures, attributing the failure to methodological problems in the test. They contend that these tests were designed to test heterogeneous groups, while military aviation candidates tend to be a homogeneous group. Also, given that the training is voluntary, it may only attract certain personality types.

Despite a lack of success in this area, the Navy and Air Force continue to investigate the use of computer-based personality measurement tests in the pilot selection process. Helton(1993) evaluated a pilot personality questionnaire for the Navy and found that one of 12 scales, assertiveness, correlated significantly with pass/fail in pilot training.

D. JOB SAMPLE MEASURES

A job sample test is an artificially created situation where the individual being tested is required to perform a range of the functions that would actually be required on the job(Hunter and Burke, 1995). Until the early 1970's, job sample tests for pilots were usually conducted in light, single engine aircraft. Boyle and Hagin(1953),

Bigbee(1954), and Ambler and Wallace(1967) conducted studies of pilot candidates who received light plane instruction prior to entering the training pipeline. Each of these studies found a correlation between the light plane instruction and later performance in flight training. The Air Force continues to use light plane evaluation measures in its pilot selection while the Navy has phased them out due to budgetary constraints.

As computers become more powerful and accessible, less expensive simulators are being used in job sample measures. Hill and Goebel(1971) developed a simulator-based job sample measure using a Link General Aviation Trainer(GAT-1). This device underwent a series of modifications and eventually became the Automated Pilot Aptitude Measurement System(APAMS). Hunter and Thompson(1978) used the APAMS to conduct two studies of pilot candidates for which they reported a correlation between APAMS and later flight performance. In the civilian sector, Stead(1991) reported on the validity of a simulator check-ride conducted by Qantas in the B747. He found that performance on the simulator check-ride correlated significantly with performance on all training criteria.

E. INTEGRATING THE RESEARCH FINDINGS

A technique that allows the integration of findings from multiple research studies and produces a single correlation that efficiently describes the validity results from a number of studies is called meta-analyis. Hunter and Burke(1994) have applied this technique to studies of pilot selection. Their results are presented in Table 3-1. As can be seen from the table, job sample measures were found to be among the best predictors, followed closely by gross dexterity. In interpreting their results, Hunter and Burke make the following conclusions. First, if the lower 95% confidence limit is positive, this leads to the conclusion that the mean sample weighted correlation across studies is assumed to be non-zero but may be influenced by moderator variables. Second, if the lower 95% confidence limit is negative while the upper 95% confidence limit is positive, this leads to uncertainty about the whether the true value of the mean sample weighted correlation is other than zero.

Table 3-1
Meta-analysis Results

PREDICTOR	r_{mean}	N_x	L_{95}	U_{95}
General ability	0.13	14	-0.05	0.30
Verbal Ability	0.12	17	-0.09	0.33
Quant. Ability	0.11	34	0.01	0.21
Spatial Ability	0.19	37	0.05	0.32
Mechanical	0.29	36	0.11	0.48
General Info.	0.25	13	0.06	0.44
Aviation Info.	0.22	23	0.06	0.38
Gross Dexterity	0.32	60	0.15	0.49
Fine Dexterity	0.10	12	-0.09	0.29
Perceptual Speed	0.20	41	0.05	0.35
Reaction Time	0.28	7	0.16	0.39
BioInventory	0.27	21	0.07	0.47
Age	-0.10	9	-0.25	0.05
Education	0.06	9	-0.16	0.27
Job Sample	0.34	16	0.19	0.55
Personality	0.10	46	-0.16	0.37

r_{mean} Mean sample weighted correlation
N_x Number of studies
L_{95} Lower 95% confidence interval
U_{95} Upper 95% confidence interval

Pilot selection has been and continues to be of great interest to both military and civilian researchers. Despite over 70 years of research, an optimal selection method has yet to evolve. While there have been significant advances in the selection measures used since their inception during World War I, only a little over half of the variance in pilot performance can be explained by the current measures.

General cognitive tests continue to be the foundation of both the Navy and Air Force selection processes because they provide the best "bang for the buck." It has been suggested however, that because there has been very little improvement in the validity of these measures since World War II, they have come about as far as they can as a prediction measure. Ease of administration and the relatively low cost per individual are the primary reasons for the continued dominance of this type of assessment.

The answer to improved validity in the pilot selection process most likely lies in the areas of psychomotor and job sample measures. The rapid increase in technology that we have experienced in the last few years has led to the development of incredibly realistic computer-based simulators. These simulators, only dreamed of a decade ago, should provide military researchers an ideal platform to design valid psychomotor and job sample measures. These

simulators are not inexpensive to build and operate, but given the escalating costs of attrition from military flight training, even a marginal improvement in the attrition rate could lead to substantial savings. These savings combined with the falling cost of computer technology may soon make these measures cost effective.

IV. CONSTRUCTION OF PROBABILITY MODEL

A. DATA SOURCES

The data set for this analysis was obtained from three separate sources. Data on flight school performance was obtained from the Naval Operational Medical Institute(NOMI) located in Pensacola, Florida. These data included flight performance scores, Aviation Selection Test Battery scores, attrition data, academic performance, and dates of entry and completion of the flight program.

Demographic and Naval Academy admissions data was obtained from records provided by the Naval Personal Research and Development Center(NPRDC) in San Diego, California. And finally, data on Academy performance including academic and military quality point ratings, Order of Merit, and academic major was obtained from the Naval Academy Office of Institutional Research in Annapolis, Maryland. These data sources were matched by social security number and verified by midshipman identification number.

The database that resulted from this merge contained observations on 2,340 Naval Academy midshipmen from the classes of 1995 and 1996. Several constraints were placed on the original data set in order to isolate those

observations to be included in the study. First, those individuals who did not graduate were removed, leaving 1765 cases. Second, those individuals who were not selected for aviation training were excluded. This reduced the data set to 711 observations. Third, due to the inherent differences in the training curriculum, Naval Flight Officers were eliminated. This reduced the data set to 495 observations. Fourth, student pilots who were still in the training pipeline, which resulted in incomplete grades, were eliminated. And finally, student pilots who were immediately disqualified from training during their initial flight physical were removed. This left a total of 272 observations in the data set for this analysis.

B. ANALYSIS OF VARIABLES

For this analysis a two-stage Heckman procedure is utilized requiring both a probit and ordinary least-squares model. The variables for these models were analyzed as follows:

1. Dependant Variables

The dependent variable used in the first-stage probit model was **ADVANCE**. This variable takes on a value of zero for those individuals who failed to complete either the aviation indoctrination or primary flight training phases and a value of 1 for those individuals who completed through

the primary flight training stage and moved on to advanced training. The data indicate that 87.1% of Naval Academy graduates completed the primary stage of flight training while 12.9% attrited from the program.

The dependent variable for the second-stage ordinary least-squares model was **FLIGHTGRD**. This is the cumulative average score for the 54 sorties that comprise the primary flight training syllabus. This score is derived from the flight instructor rating based on a 4.0 scale. Each maneuver performed during a given sortie is graded as one of the following: above average(4.0), average(3.0), below average(2.0), or unsatisfactory(0.0). The distribution of these scores is displayed below in Figure 4-1.

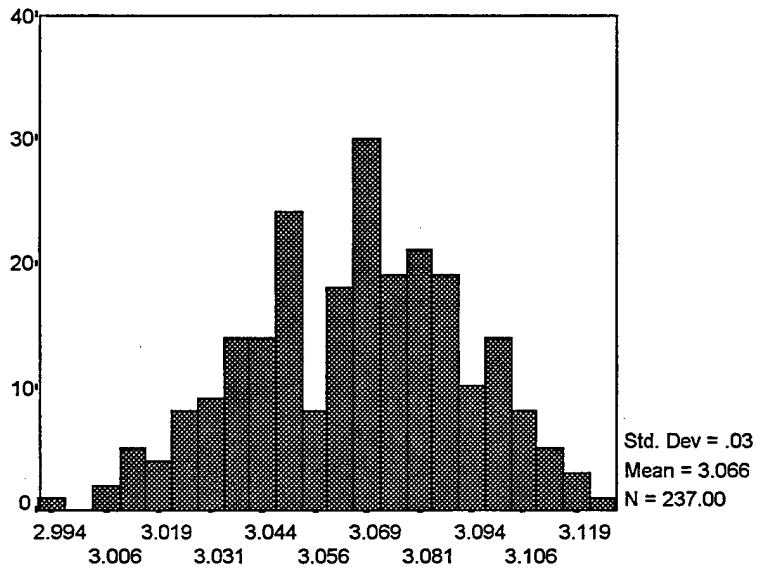

Figure 4-1 Flight Grade Distribution

2. Independent Variables

The selection of independent variables was determined from previous research. Descriptive statistics for these variables are displayed in Table 4-1. Each of these variables and their expected impact on the models is discussed below. Dummy variables were created to represent minority status, marital status, gender, previous flight experience, academic major, change in major, and service.

MINORITY- This variable was created to indicate minority status. Based on demographic data from NPRDC, the majority of individuals in the study, 91.2%, are Caucasian and are coded as zero. Due to the relatively small representation of individual minority groups, all individuals that did not fall into the majority group were combined into a single minority group. This group represented 8.8% of the data and are coded as 1. Previous research has shown that a smaller percentage of minority candidates meet the minimum standards of the selection tests. For this reason the expected sign of this coefficient is negative.

MARRIED- This variable indicates marital status and takes on a value of 1 if the individual is married and zero otherwise. Because being married places extra demands upon the student pilot, the expected sign is negative.

Table 4-1 Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
ADVANCE	272	.00	1.00	.8713	.3355
FLTGRADE	237	2.991	3.124	3.06584	2.58E-02
AGE	272	21.50	27.50	23.3971	1.1062
AQPR	272	208.00	396.00	292.4963	42.2504
MQPR	272	268.00	387.00	332.6471	25.5421
SATM	272	471.00	790.00	631.5000	63.2247
TIS	272	311.00	714.00	511.5588	92.7015
AQR	272	3	9	6.46	1.46
BIPILOT	272	4	9	7.61	1.32
PFAR	272	4	9	6.16	1.43
DELTAMAJ	272	.00	1.00	5.5E-02	.2287
GENDER	272	.00	1.00	7.4E-02	.2615
MAJOR	272	.00	1.00	.6838	.4658
MARINE	272	.00	1.00	.1397	.3473
MARRIED	272	.00	1.00	2.2E-02	.1471
MINORITY	272	.00	1.00	8.8E-02	.2842
PREFLY	237	.00	1.00	5.5E-02	.2282
Valid N (listwise)	237				

Note: The notation E represents the operation of taking a numerical value to a particular power. For example 3.14E-02 equals 3.14×10^{-2} or 0.0314.

GENDER- This variable was created to separate males from females. Males, who represent 92.6% of the sample, are coded zero. Females, representing 7.4%, are coded 1. Selection tests and performance evaluations are designed to be gender neutral so it is unclear what the sign of this coefficient will be.

PREFLY- This variable is created to separate those individuals who had previous flight experience prior to entry into flight training. Information for this variable is self reported by the student aviators. Based on personal experience, it is suggested that a minimum threshold of flight experience must be obtained before this experience becomes meaningful. An arbitrary value of 20 hours was selected as the threshold for this variable. Individuals with less than 20 hours of experience are coded as zero. Those with more than 20 hours are coded as one. The expected sign for this coefficient is positive.

MAJOR- This variable separates those individuals who graduated with a technical degree, a so-called group 1 or group 2 major at the Academy, from those who graduated with a non-technical degree. Technical majors are coded as one and non-technical majors as zero. Technical degrees involve a higher level of math and engineering comprehension than non-technical majors. For this reason the expected sign for

this coefficient is positive.

DELTAMAJOR- This variable identifies those individuals who changed from a technical major to a non-technical major. This change would indicate that these individuals were having difficulty with the more demanding levels of math and engineering associated with a technical degree. The expected sign on this coefficient is negative.

MARINE- This variable is created to distinguish Marine aviators from naval aviators. Marine aviators must attend The Basic School prior to reporting to Pensacola. There they learn tactics and basic Marine doctrine. It is suggested that the additional military experience that these individuals receive will make them less likely to attrite from training than their Navy counterparts. For this reason the expected sign for this coefficient is positive.

AGE- This variable is the student naval aviator's age at the beginning of flight training. This variable was derived by subtracting as student's date of birth from their Aviation Indoctrination class convening date. Previous research indicates that age has a negative influence during flight training. The expected sign on this coefficient is negative.

BIPILOT- This score is derived from the biographical information section of Aviation Selection Test Battery

(ASTB). The raw score is converted into a standard nine-point scale. This portion of the test is specifically designed to predict attrition from flight training. The higher the score, the less likely an individual is to attrite. For this reason the expected sign on this coefficient is positive.

PFAR- This score is the pilot flight aptitude from the ASTB and is designed to predict basic flight performance. Based on a standard nine-point scale, a higher score should reflect better flight performance. The expected sign on this coefficient is positive.

AQPR- This score is the individual's academic quality point rating or grade point average for academic performance while at the Naval Academy. The expected sign on this coefficient is positive.

MQRP- This variable represents the individual's military quality point rating, a measure of military performance while at the Naval Academy. The expected sign this coefficient is positive.

TIS- This variable is the score from the Strong-Campbell Technical Interest Survey taken as part of the Naval Academy admissions process. The expected sign on this coefficient is positive.

SATM- This variable is the average score achieved on

the math section of the Scholastic Aptitude Test reported during the Naval Academy admissions process. A higher score indicates higher mathematical ability. The expected sign on this coefficient is positive.

AQR- This variable is the individual's academic quality rating from the ASTB. Based on a nine point standard scale, this score is designed to predict academic performance during flight training. The expected sign on this coefficient is positive.

C. MODEL DESCRIPTION

The model specification used to describe the determinants of a Naval Academy graduate's flight grades takes the following form:

$$F_i = f(g_i, p_i, t_i, s_i, z_i) + u_i. \qquad 4\text{-}1$$

F_i is the flight grade of the *i*th individual, g_i is a measure of cognitive ability, p_i a measure of psychomotor ability, t_i personality measures, s_i job sample measures, z_i represents other factors affecting performance such as race an gender, and u_i is a random disturbance term reflecting unobserved ability characteristics and the measurement error of performance statistics.

It is assumed that this function will take the form of the linear equation $Y_i=a+bX_i+\epsilon_i$, where Y_i, the dependent variable, represents flight grades and the right hand X_i

term represents a vector of independent variables. Usually, when trying to explain or predict changes in Y_i due to X_i, ordinary least-squares, which minimizes the vertical sum of the squared deviation from the fitted line, is used. The ordinary least-squares procedure assumes the following:

- The relationship between Y and X is linear.
- The X_i's are nonstochastic variables whose values are fixed.
- The error term has zero expected value and constant variance for all observations.
- The random variables ϵ_i are statistically independent.
- The error term is normally distributed.

When these assumptions are satisfied, estimators are unbiased and consistent. The normality assumption permits hypothesis testing to be conducted when the sample size is relatively small.

1. Selectivity Bias

The term bias refers to the potential misestimate of the effect of a treatment on the outcome. Suppose for example that one is studying the wages of women or the automobile purchasing behavior of a random sample of the population. It is possible to know the actual wages of those women who are working, but it is not possible to know the reservation wage for those who are not. In the automobile example, for those who happened to purchase a

car, the expenditure can be recorded. But for both those who did and did not purchase a car, one cannot measure the maximum amount they would have been willing to pay for the automobile.

In both of the preceding examples, the dependent variable is censored, that is, information is missing for the dependent variable, but the corresponding information for the independent variables is present. This situation is referred to as selectivity bias. A similar situation exists in our model of flight grades in that flight grades exist only for those who were able to complete the program. It can be shown that ordinary least-squares estimation of the censored regression model may fail to satisfy the assumption that the error term of the model has zero expected value and, if this is so, would generate biased and inconsistent parameter estimates. In order to correct for the selectivity bias, however, a relatively simple two-stage estimation procedure that yields consistent parameter estimates can be employed.

2. Censored Regression Model

This section will show how the ordinary least-squares estimation of a censored regression model can generate biased and inconsistent parameter estimates. We will then describe how to correct for this situation. As mentioned

before, in the censored regression model information for some observations of the dependent variable are missing while corresponding information for the independent variables is present.

Suppose that the underlying continuous version of the model is given by:

$$\mathbf{F_i} = \beta_0+\beta_1\mathbf{X_{1i}}+\beta_2\mathbf{Z_i}+\upsilon_{Fi}, \qquad \mathbf{4-2}$$

where F_i represents the final flight grade for those individuals who completed the primary stage of flight training. For those individuals who did not complete, this value equals zero. Assume that there is a survival equation that must be equal to or greater than some minimum score (C) for F_i to be observed. The survival equation is a function of flight performance(F_i), academic performance(A_i), and military performance(M_i) and takes the form of:

$$\mathbf{Y_{Si}=aF_i+bA_i+cM_i}+\upsilon_i, \qquad \mathbf{4-3}$$

where[1]

$$F_i=\beta_0+\beta_1X_{1i}+\beta_2Z_i+\upsilon_{Fi},$$

$$A_i=\alpha_0+\alpha_1X_{2i}+\alpha_2Z_i+\upsilon_{Ai}, \text{ and}$$

$$M_i=\gamma_0+\gamma_1X_{3i}+\gamma_2Z_i+\upsilon_{Mi}.$$

In these equations the X_{ji}(j=1,2,3) variables represent a vector of variables that are unique to the individual

[1]In reality each of these equations may have its own minimum score.

equation. The Z_i variables represent a vector of variables common to the three equations.

Let

$$(\beta_0 + \alpha_0 + \gamma_0 - C) = \theta_0,$$

$$(\beta_2 + \alpha_2 + \gamma_2) = \theta_1, \text{ and}$$

$$(\upsilon_{Fi} + \upsilon_{Ai} + \upsilon_{Mi} + \upsilon_i) = \epsilon_i.$$

Equation 4-3 can now be expressed as follows:

$$\text{if} \quad \mathbf{Y_{si}} = \theta + \beta_1 \mathbf{X_{1i}} + \alpha_1 \mathbf{X_2} + \gamma_1 \mathbf{X_{3i}} + \theta_1 \mathbf{Z_i} + \epsilon_i \geq 0 \text{ then } \mathbf{d_i = 1}$$

$$\text{if} \quad \mathbf{Y_{si}} = \theta + \beta_1 \mathbf{X_{1i}} + \alpha_1 \mathbf{X_2} + \gamma_1 \mathbf{X_{3i}} + \theta_1 \mathbf{Z_i} + \epsilon_i < 0 \text{ then } \mathbf{d_i = 0}. \qquad \mathbf{4-4}$$

This implies that

$$\mathbf{P(d_i = 1) = P(advance) = F(}\theta + \beta_1 \mathbf{X_{1i}} + \alpha_1 \mathbf{X_{2i}} + \gamma_1 \mathbf{X_{3i}} + \theta_1 \mathbf{Z_i}\mathbf{)}, \quad \mathbf{4-5}$$

where F is a cumulative normal probability function.

Returning to equation 4-2 and evaluating given the sample selection rule in equation 4-4,

$$E(F_i | X_{1i}, Z_i, Y_{Si} - C \geq 0) = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i + E(\upsilon Fi | X_{1i}, Z_i, Y_{Si} - C \geq 0).$$

James Heckman shows that[2]

$$\mathbf{E(}\upsilon_{Fi} | \mathbf{X_{1i}, Z_i, Y_{Si} - C \geq 0) = (}\sigma_{FYs} / \sigma_{Ys}\mathbf{)}\lambda_i,^{3} \qquad \mathbf{4-6}$$

where σ_{FYs} is the covariance$(\upsilon_F \upsilon_{Ys})$ and σ_{Ys} is the standard deviation(Y_S).

---

[2]Heckman, James J., "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Variables and a Simple Estimator for Such Models", *Annals of Economic and Social Measurement*, 1976.

[3]Each of the individual equations used to create equation 4-3 would require a probit equation to calculate an individual λ term. For modeling purposes these terms are summed into an aggregate λ.

Equation 4-2 can now be written as

$$F_i=\beta_0+\beta_1 X_{1i}+\beta_2 Z_i+(\sigma_{FYs}/\sigma_{Ys})\lambda_i+\mu_i. \qquad 4\text{-}7$$

Heckman has shown that

$$\lambda_i = f(\theta+\beta_1 X_{1i}+\alpha_1 X_2+\gamma_1 X_{3i}+\theta_1 Z_i)/ F(\theta+\beta_1 X_{1i}+\alpha_1 X_2+\gamma_1 X_{3i}+\theta_1 Z_i), \qquad 4\text{-}8$$

where *f* is the probability density function of a standardized normal variable, and F is the corresponding cumulative distribution function. If we have estimates of λ_i, we can use them to adjust for the fact that the expected value of the censored error term shown in equation 4-6 may not equal to zero. The Heckman two-stage process is used to yield consistent estimates of β_0, β_1, and β_2.

D. TWO-STAGE HECKMAN PROCEDURE

In the first-stage, λ_i is estimated utilizing the probit model:

$$P_i=F(\theta+\beta_1 X_{1i}+\alpha_1 X_2+\gamma_1 X_{3i}+\theta_1 Z_i), \qquad 4\text{-}9$$

where P_i equals the probability of completing primary flight training. This probit model is estimated with a maximum-likelihood technique by distinguishing those observations for which $F_i>0$ from those for which $F_i=0$. From the estimated parameters of the probit model, $\hat{\lambda}_i$ may be calculated using equation 4-8.

The second stage of the two-stage estimator uses the

following ordinary least-squares model:

$$\mathbf{F}_i=\beta_0+\beta_1\mathbf{X}_i+\beta_2\mathbf{Z}_i+\sigma\hat{\lambda}_i+\nu_i, \qquad \textbf{4-10}$$

where $\hat{\lambda}_i$ has been added as an additional explanatory variable and ν_i is an error term with expected value equal to zero in large samples. Because $\hat{\lambda}_i$ approaches λ_i as the sample size gets large, and λ_i normalizes the mean υ_{Fi} to zero, ordinary least-squares estimation of equation 4-10 yields consistent estimates of β_0, β_1, and β_2.

V. ANALYSIS OF EMPIRICAL MODELS

This section contains the analysis of the multivariate models discussed in the previous chapter. Using the Heckman two-stage regression technique which combines the multivariate probit model with the (adjusted) ordinary least-squares(OLS) regression model we examine the individual effects of the independent variables on flight performance holding several other variables constant. The first-stage probit model is designed to create a new independent variable, $\hat{\lambda}$, which is included in the second-stage OLS model as an explanatory variable to correct for potential selectivity bias.

During the analysis, two specifications, an initial and final specification, were created for both parts of the two-stage procedure. In the initial specification all relevant independent variables are included. In the alternate specification, the first-stage probit model is altered to include only those variables with a "*t*"-statistic[4] greater than one. Also retained are the gender and minority variables which have *t*-statistics less than one but are of interest to the study. The second-stage alternate

[4]The calculated "*t*" is equal to the estimated coefficient divided by the standard error. This value approaches the *t*-statistic in large samples.

specification also retained only those variables with a *t*-statistic greater than one and a few select variables of interest to the study. Additionally, the OLS model was run twice for the final specification. In the first estimation, the model included the $\hat{\lambda}_i$ independent variable. In the second estimation, this variable is eliminated to display the results without an adjustment for selectivity bias.

A. FIRST-STAGE PROBIT MODEL

The initial specification for the first-stage model is as follows:

$$\text{Prob(advance)} = F(\beta_0 + \beta_1\text{BIPILOT} + \beta_2\text{PFAR} + \beta_3\text{AQPR} + \beta_4\text{DELTAMAJOR} + \beta_5\text{MAJOR} + \beta_6\text{MARINE} + \beta_7\text{AGE} + \beta_8\text{GENDER} + \beta_9\text{MARRIED} + \beta_{10}\text{MINORITY} + \beta_{11}\text{MPQR} + \beta_{12}\text{AQR} + \beta_{13}\text{TIS} + \beta_{14}\text{SATM})$$

where F is the cumulative normal probability function of the underlying random process.

The results for the initial probit specification are displayed in Table 5-1.

Table 5-1
Probit Analysis of Likelihood of Naval Academy Graduates Completing Primary Stage of Flight Training: Initial Specification

INDEPENDENT VARIABLE	COEFFICIENT ESTIMATE	T-STATISTIC
BIPILOT	0.21577	2.91967***
PFAR	0.13094	0.90455
AQPR	0.01064	2.54256**
DELTAMAJOR	-0.53254	-2.42007**
MAJOR	0.57910	2.30031**
MARINE	0.80868	1.85285*
AGE	0.7483	0.65925
GENDER	0.2715	0.61111
MARRIED	-0.74831	-1.13623
MINORITY	-0.26249	-0.75288
MPQR	-0.00451	-0.73788
AQR	-0.09065	-0.62157
TIS	-0.00155	-1.15827
SATM	-0.00273	-1.15185
INTERCEPT	-3.78041	-1.00691
SAMPLE SIZE	272	

CHI-SQUARE (df=258)= 289.090, significance =.089
*** SIGNIFICANT AT 0.01 LEVEL
** SIGNIFICANT AT 0.05 LEVEL
* SIGNIFICANT AT 0.1 LEVEL

Many of the variables in Table 5-1 are not statistically significant. However, the chi-square statistic is significant which indicates that the model is a fairly good fit. The variables BIPILOT, AQPR, DELTAMAJOR, MAJOR, and MARINE are all significant with coefficients that match their predicted sign. The AGE variable, which was expected to have a negative coefficient, has the opposite sign in this model but is not statistically significant.

The alternate specification was developed by taking the significant variables in the initial specification as a base model. Then to allow for the possibility of multicolinearity in the original specification, the other variables were individually added back into the model and the change in the chi-square statistic caused by the addition of that variable was analyzed. Variables that raised the value of the chi-square statistic were retained in the model, while those that reduced the value were omitted. The goal was to build the best possible first-stage model giving proper account to theory, multicolinearity, and tests of significance. The regression results for the alternate specification are shown in Table 5-2.

Table 5-2
Probit Analysis of Likelihood of Naval Academy Graduates Completing Primary Stage of Flight Training: Alternate Specification

INDEPENDENT VARIABLE	COEFFICIENT ESTIMATE	T-STATISTIC
BIPILOT	0.20461	2.91743***
AQPR	0.00839	2.72541***
DELTAMAJOR	-0.51660	-2.42801**
MAJOR	0.43444	1.96295**
MARINE	0.72423	1.82212*
AGE	0.08938	0.84737
GENDER	0.24898	0.61072
MINORITY	-0.23999	-0.70480
INTERCEPT	-5.30902	-1.72949
SAMPLE SIZE	272	

CHI-SQUARE (df=263)= 393.595, significance = .001
*** SIGNIFICANT AT 0.01 LEVEL
** SIGNIFICANT AT 0.05 LEVEL
* SIGNIFICANT AT 0.1 LEVEL

All variables in the alternate specification retain the same sign as in the initial specification. Additionally, the significance level of all variables, with the exception of AQPR, did not change. The significance level of AQPR increased to the 0.01 level in the alternate specification. The goodness of fit chi-square statistic shows an increase in the alternate specification, rising from 289.090 (df=258) in the initial specification to 393.595 (df=263) in the alternate specification. The decrease in p-value associated with this change in chi-square is from 0.089 to 0.001[5]

[5]The p-value is the probability of obtaining a chi-square value at least as large as the calculated value, when the null hypothesis is true.

1. Goodness of Fit

An important measure of quality of the individual probit equations is the Pearson goodness of fit chi-square statistic. The chi-square statistic from each probit equation was compared against the critical value. This statistic accepts the null hypothesis that the explanatory variables in the model do not contribute to the overall equation if the computed significance level does not exceed the critical value. In the initial specification the chi-square value was large enough to reject the null hypothesis at the 0.089 level and in the alternate specification the chi-square value was large enough to reject the null hypothesis at the 0.001 level. The chi-square value in the alternate specification indicates that the variables retained in the alternate specification contribute more to the quality of the equation than those in the initial specification.

2. Notional Pilot

Although the primary purpose of the first-stage probit model was to generate the $\hat{\lambda}_i$ variable in order to estimate consistent and unbiased results in the second-stage OLS model, it is interesting to interpret the results of the probit regressions. The coefficients of the variables are converted to an estimate of the change in probability of

completing primary flight training for a notional pilot who possesses a given set of characteristics. The notional pilot was defined by setting all the dummy variables to zero and by using the average value of the continuous variables. Once the "base-case" pilot was defined, each variable was manipulated individually to generate the change in probability of completing primary flight training associated with a change in that independent variable[6]. The notional pilot has the following characteristics:

- Navy
- Caucasian
- Male
- Non-technical degree
- Did not change out of a technical major
- AQPR of 2.92

The change in probability associated with changes in these characteristics are displayed in Table 5-3. The change in AQPR from 2.92 to 3.02 was chosen to illustrate the change associated with a 0.1 change in AQPR.

[6]This is an approximation based on the method found in Pindyck and Rubinfeld.

Table 5-3 Change in Probability

Variable	Initial Value	New Value	Delta Prob.
AQPR	2.92	3.02	Increase 3.4%
MARINE	No	Yes	Increase 15%
MAJOR	Non-technical	Technical	Increase 11%
DELTAMAJOR	No	Yes	Decrease 18%
BIPILOT	5	6	Increase 16%

B. SECOND-STAGE OLS MODEL

The second-stage of the Hecknam regression technique is the construction of an ordinary least-squares model that includes the $\hat{\lambda}_i$ as an independent explanatory variable to correct for potential selectivity bias. The initial specification of this model takes the following form:

$$F_i = \beta_0 + \beta_1 BIPILOT + \beta_2 PFAR + \beta_3 AQPR + \beta_4 DELTAMAJOR + \beta_5 MAJOR + \beta_6 MARINE + \beta_7 AGE + \beta_8 GENDER + \beta_9 MARRIED + \beta_{10} MINORITY + \beta_{11} MPQR + \beta_{12} AQR + \beta_{13} TIS + \beta_{14} SATM + \beta_{15} PREFLY + \beta_{16} \hat{\lambda}_i + \nu_i .$$

The regression results for this model are displayed in Table 5-4.

Table 5-4
Regression Results for Naval Academy Graduate's Flight Grades: Initial Specification

DEPENDENT VARIABLE:FLIGHT GRADES

INDEPENDENT VARIABLE	COEFFICIENT ESTIMATE	T-STATISTIC
BIPILOT	-0.097	-0.0691
PFAR	0.570	4.9590***
AQPR	0.012	2.2957**
DELTAMAJOR	0.296	0.7328
MAJOR	0.17	0.4832
MARINE	-0.154	-0.2934
AGE	-0.136	-0.7603
GENDER	-0.614	-1.0674
MARRIED	1.10	1.0043
MINORITY	-0.241	-0.4373
MQPR	0.012	1.6161*
AQR	-0.0016	-0.7570
TIS	0.0024	1.4930*
SATM	0.0032	0.1132
PREFLY	1.790	2.7221***
LAMDA	-1.182	-1.0082
INTERCEPT	299.34	41.531
SAMPLE SIZE	237	
R-SQUARE	0.378	
ADJ R-SQUARE	0.336	
F-STATISTIC	8.962	

*** SIGNIFICANT AT 0.01 LEVEL
** SIGNIFICANT AT 0.05 LEVEL
* SIGNIFICANT AT 0.1 LEVEL

As can be seen from the regression results, the variables PFAR, AQPR, PREFLY, TIS, and MQPR are statistically significant in this model. Several variables that were significant in the first-stage probit model, MARINE, DELTAMAJOR, and MAJOR are not statitically significant in the second-stage OLS model. This is consistent with the theory used in formulating equation 4-3, that the overall survival equation is the sum of multiple equations acting simultaneously and that some of the variables are only germaine to individual parts of the overall equation.

In the interest of parsimony, only those variables significant at the 0.10 level, along with the MINORITY, GENDER, and $\hat{\lambda}_i$ variables are retained for the alternate specification. The alternate specification is as follows:

$$F_i = \beta_0 + \beta_1 PFAR + \beta_2 APQR + \beta_3 GENDER + \beta_4 MINORITY + \beta_5 MARRIED + \beta_6 MQPR + \beta_7 TIS + \beta_8 PREFLY + \beta_9 \hat{\lambda}_i + \nu_i.$$

The regression results for this model are displayed in Table 5-5. The $\hat{\lambda}_i$ variable is not significant in the alternate specification and there is actually a decrease in the *t*-statistic in the alternate model. The non-significance of this variable leads to the conclusion that selectivity bias is not pervasive for this sample of flight

students. A possible explanation is that, while the data is censored, the covariance between the survival equation and the flight grade equation is fairly small. If this is the case, the OLS model without the $\hat{\lambda}_i$ variable is a more appropriate model for this sample. A comparison of the alternate specification with and without the $\hat{\lambda}_i$ variable shows that while there are minor changes in the *t*-statistics, none of the variables change significance levels and the signs on the coefficients remain the same.

As a result of the non-significant $\hat{\lambda}_i$, the alternate specification is run without the correction for selectivity bias. The results for this regression are displayed in Table 5-6. For ease of data interpretation, the Beta coefficient is included in Tables 5-5 and 5-6. The Beta coefficient is a normalized coefficient that can be used to assess the relative importance of the independent variables in the model.

Table 5-5
Regression Results for Naval Academy Graduate's Flight Grades: Alternate Specification

DEPENDENT VARIABLE: FLIGHT GRADES

INDEPENDENT VARIABLE	COEFFICIENT ESTIMATE	T-STATISTIC	BETA
PFAR	0.580	5.529***	0.331
AQPR	0.014	3.139***	0.219
GENDER	-0.461	-0.872	-0.047
MARRIED	0.914	0.930	0.051
MINORITY	-0.357	- 0.683	-0.038
MQPR	0.012	1.696*	0.117
TIS	0.0028	1.880*	0.101
PREFLY	1.683	2.706**	0.149
LAMDA	-0.977	-0.966	-0.055
INTERCEPT	294.535	123.349	
SAMPLE SIZE	237		
R-SQUARE	0.375		
ADJ R-SQUARE	0.350		
F-STATISTIC	15.129		

*** SIGNIFICANT AT 0.01 LEVEL
** SIGNIFICANT AT 0.05 LEVEL
* SIGNIFICANT AT 0.1 LEVEL

Table 5-6
Ordinary Least-squares Regression Results for Naval Academy Graduate's Flight Grades: Alternate Specification Without Sample Selection Correction

DEPENDENT VARIABLE: FLIGHT GRADES

INDEPENDENT VARIABLE	COEFFICIENT ESTIMATE	T-STATISTIC	BETA
PFAR	0.603	5.851***	0.342
AQPR	0.013	3.092***	0.215
GENDER	-0.425	-0.805	-0.044
MARRIED	0.748	0.773	0.042
MINORITY	-0.468	-0.916	-0.049
MQPR	0.013	1.817*	0.125
TIS	0.0028	1.843*	0.099
PREFLY	1.657	2.666**	0.146
INTERCEPT	284.535	125.759	
SAMPLE SIZE	237		
R-SQUARE	0.372		
ADJ R-SQUARE	0.350		
F-STATISTIC	16.908		

*** SIGNIFICANT AT 0.01 LEVEL
** SIGNIFICANT AT 0.05 LEVEL
* SIGNIFICANT AT 0.1 LEVEL

The regression results of the alternate specification in Table 5-5 show that the coefficient sign for all variables remains the same as in the initial specification. The PFAR variable is significant at the 0.01 level and shows that a higher score on the pilot flight aptitude rating of the Aviation Selection Test Battery predicts higher flight grades. The Beta coefficient indicates that an increase of one standard deviation in the PFAR score would result in a 0.342 standard deviation increase in the final flight grade.

Another variable significant at the 0.01 level is AQPR. Table 5-5 shows that a higher academic quality point rating while at the Academy predicts higher flight grades. The Beta coefficient indicates that a one standard deviation increase in AQPR leads to a 0.215 standard deviation increase in the final flight grade. PREFLY is significant at the 0.05 level showing that previous flight experience is a significant predictor of higher final flight grades. MQPR is marginally significant at the 0.1 level showing that a higher military quality point rating predicts higher flight grades. TIS is also marginally significant at the 0.1 level indicating that a higher score on the Strong-Campbell Technical Interest Survey predicts higher flight grades. The remaining variables, including GENDER and MINORITY, contained in the alternate specification were not statistically significant.

In the alternate specification the variables PFAR, AQPR, MPQR, and TIS each show incremental improvement in their *t*-statistic, while PREFLY shows a modest decrease but remains significant at the 0.05 level. The adjusted R-squared for the model, which explains the amount of variance accounted for by the model, increased from 0.336 in the initial specification to 0.35 in the alternate specification. This leads to the conclusion that the

alternate specification does a somewhat better job of explaining the variance in the data when the degrees of freedom of the model are taken into account.

VI. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

This thesis investigates the relationship between observable characteristics and performance in primary flight training using a two-stage Heckman regression procedure. The hypotheses identified in the introduction will now be restated and compared with the results of this procedure.

Hypothesis 1.

- Naval Academy graduates who score higher on the biographical inventory are more likely to complete primary flight training than those with lower scores.

Based on the significance level of the BIPILOT variable in the probit model, the null hypothesis that the coefficient on this variable is zero is rejected. Rejection of the null hypothesis supports the alternate hypothesis that the true value of this coefficient is not zero. This result indicates that there is a relationship between higher scores on the biographical inventory and successful completion of primary flight training.

Hypothesis 2.

- Naval Academy graduates who score higher on the pilot flight aptitude rating achieve higher flight grades that those with lower scores.

The significance of the PFAR variable in the second-stage OLS equation leads us to reject the null hypothesis that the value of this coefficient is zero in favor of the alternate hypothesis that the true value of this coefficient is not zero. This indicates that there is a relationship between higher scores on the flight aptitude rating and flight grades.

Hypothesis 3.

- Naval Academy graduates with higher academic quality point ratings achieve higher flight grades than those with lower ratings.

The significance of the AQPR variable in the second-stage OLS model allows the null hypothesis that the value of this coefficient is zero to be rejected in favor of the alternate hypothesis that the value is not zero. This indicates that there is a direct relationship between the level of academic achievement at the academy and primary flight grades.

Hypothesis 4.

- Naval Academy graduates with higher military quality point ratings achieve higher flight grades than those with lower ratings.

The significance of the MPQR variable is inconclusive. The null hypothesis that there is no relationship is rejected at the 0.1 level but not the 0.05 level.

Hypothesis 5.

- Naval Academy graduates with technical degrees achieve higher scores that graduates with non-technical degrees.

The results of the probit model indicate that graduates with technical degrees are more likely to finish the primary stage of training. The null hypothesis that the coefficient of this variable in the second stage OLS is equal to zero, however, cannot be rejected. This provides evidence that the type of degree obtained from the academy is not related to primary flight grades.

Hypothesis 6.

- Male graduates achieve higher flight grades than female graduates.

Based on the significance level of the GENDER variable, the null hypothesis that the value of the coefficient is different than zero cannot be rejected. This indicates that there is no difference associated with gender in primary flight grades.

Hypothesis 7.

- Ethnic majority graduates achieve higher flight grades than minority graduates.

The null hypothesis that the value of the MINORITY coefficient is different from zero cannot be rejected. This indicates that there is no difference in primary flight

grades associated with the race of the individual.

Hypothesis 8.

- Naval Academy graduates with previous flight experience achieve higher flight grades than graduates with no previous flight experience.

The value of the PREFLY variable is significant at the 0.05 level allowing us to reject the null hypothesis that the coefficient is zero. Rejection of the null hypothesis supports the alternate hypothesis that the value of this coefficient is not zero, indicating that primary flight grades are positively influenced by previous flight experience.

Based on the results of hypothesis testing, the conclusions of this study are summarized as follows:

- The biographical inventory section of the ASTB is a valid predictor of primary flight training completion.
- The pilot flight aptitude rating of the ASTB is a valid predictor of primary flight performance.
- Academic performance at the Academy is a valid predictor of primary flight performance.
- Previous flight experience is a valid predictor of primary flight performance.

These findings are consistent with the findings of the body of research contained in the literature review. An additional conclusion of this study is that selectivity bias is not a problem with this sample of flight students. There

are however several other potential sources of bias in the statistical estimates. Assignment to aviation duty is a done on a volunteer basis. It is therefore possible that there is problem with self-selection bias in the first-stage probit analysis. Individuals who have little or no interest in aviation careers or who are deterred by the significant commitment incurred by those who attend flight training can choose not to participate in the selection process. There is also a possible problem associated with the range restriction of the variables used in the models. Because some of these variables are used as selection measures, the range that the variable can assume is artificially limited. Omitted variables may also cause a problem. In addition to the variables included in the model, there may exist other relevant variables that have not been measured or for which there is no reliable proxy. Omission of these variables might lead to bias through model specification. It is hoped that through continued research in the area of pilot selection, these and other problems can eventually be addressed.

B. RECOMMENDATIONS FOR FUTURE RESEARCH

As a result of this research several interesting avenues for future research have been identified. For example, it might be interesting to study factors that

predict success for Naval Flight Officers. Do factors that predict success of pilots also predict success for Naval Flight Officers?

The literature review revealed that there has been limited success in the use of personality measures in the selection of aviators. This has been attributed to the fact that many of the student aviators are able to "fake" the test and provide the most desirable answers. A solution to this problem would be to have a test that was not associated with selection into an aviation program. The Academy has access to data of this type in the Myers-Briggs type indicator test and related variants. These tests are conducted early during the midshipmen's time at the academy and there may be a smaller probability that the midshipman will perceive his/her answer as being linked to the aviation selection process.

One of the original goals of this study was to analyze the results of the interview data obtained during the service assignment process. Unfortunately this information no longer exists for the graduating classes included in this study. This information does exist however for subsequent classes. As these graduates complete flight training, service assignment process information will be available and could be used in subsequent analysis.

Finally, as a cost savings measure, there has been much interest in trying to identify and therefore eliminate those individuals who are most likely to attrite from flight training. But what happens to those individuals who do attrite from the training program? Do these individuals serve the minimum obligation and then leave the service? Do they go on to experience successful and rewarding careers in some other community? The human resource implications of finding answers to these questions could help detailers do a better job assigning individuals to future billets A better assignment process would help ensure that the Navy receives the maximum return on any further investment in personnel capital.

REFERENCES

Ambler, R. K. & Waters, L. K. (1967). *The value of an NROTC flight indoctrination program to naval aviation training.* Pensacola, FL: Naval Aerospace Medical Center.

Bachman, R. A. (1918). "The examination of aviators." *Naval Medical Bulletin,* v.12, pp.30-41.

Berkshire, J. R. (1967). *Evaluation of several experimental aviation selection tests.* Pensacola, FL: Naval Aerospace Medical Center.

Blower, D. J. (1997) *A cost-benefit analysis of the impact of selection testing on advanced flight training.* Pensacola, Fl: Naval Aerospace Medical Research Laboratory.

Bowman, W. R. (1995). *Cost effectiveness of service academies: New evidence from Navy warfare communities.* Paper presented at the 1996 Western Economic Association Annual Conference.

Burke, E. F. (1993). "Pilot selection in NATO: An overview." *Proceedings of the 7th International symposium on Aviation Psychology.* Columbus, OH: Ohio State University.

Carretta, T. R. (1987). *Basic Attributes Test (BAT) system: A preliminary evaluation.* Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Carretta, T. R. (1993). *Pilot Candidate Selection Method: What makes it work?* Brooks Air Force Base, TX: Armstrong Lab.

Carretta, T. R. & Malcolm, J.(1996). "U.S. Air Force pilot selection tests: What is measured and what is predictive?" *Aviation, Space, and Environmental Medicine,* v.67, pp.279-83.

Carretta, T. R. & Rea, M. J. (1994). "Pilot Candidate Selection Method: Sources of validity." *International Journal of Aviation Psychology,* v.4, pp.103-117.

Cope, A. W. (1996). *Fighter pilot - A moving constant.* Paper presented at the AMP Symposium May 1996.

Dockeray, F. C. & Isaacs, S. (1921). "Psychological research in aviation in Italy, France, England, and the American Expeditionary Forces." *Journal of Comparative Psychology*, v.1, pp.115-148.

Dolgin, D. & Gibb, G. (1989). *Personality assessment in aviator selection.* Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Fiske, D. W. (1947). "Validation of naval aviation cadet selection tests against training criteria." *Journal of Applied Psychology,* v.31, pp.601-614.

Fleishman, E. A. (1956). "Psychomotor selection tests: Research and application in the United States Air Force." *Personnel Psychology*, v.9, pp.449-467.

Frank, L. H. & Baisden, A. G. (1993). "The 1992 Navy and Marine Corps Aviation Selection Test Battery Development." *Proceedings of the 35th Annual Conference of the Military Testing Association.*

Griffin, G. R. & Mosko, J. D. (1977). *Naval aviation attrition 1950-1976: Implications for the development of future research and evaluation.* Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Griffin, G. R. & North, R. A. (1977). *Aviator selection 1919-1977.* Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Helton, K. & Street, D. R. (1992). *The Five-Factor Personality Model and naval aviation candidates.* Pensacola, Fl: Naval Aerospace Medical Research Laboratory.

Henman, V. A. (1919). "Air service tests of aptitude for flying." *Journal of Applied Psychology*, v.3, pp.103-109.

Hill, J. W. & Goebel, R. A. (1971). *Development of automated GAT-1 performance measures.* Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Hilton, T. & Dolgin, D. (1991). *Pilot selection in the military of the free world.* Pensacola, FL: Naval Aerospace Medical Research Lab.

Hunter, D. R. & Burke, E. F. (1990). *An annotated bibliography of the aircrew selection literature.* Alexandria, VA: Army Research Institute.

Hunter, D. R. & Burke, E. F. (1992). *Meta-analysis of aircraft pilot selection measures.* Alexandria, VA: Army Research Institute.

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

Hunter, D. R. & Burke, E. F. (1995). *Handbook of pilot selection. Brookfield,* VT: Ashgate.

Hunter, D. R. & Thompson, N. A. (1978). *Pilot selecion system development.* Brooks Air Force Base, TX: Air Force Human Resource Laboratory.

Kinney, G. (1987). *Relationship of the Basic Attributes Test to tactical pilot performance.* Wright-Patterson Air Force Base, OH: Air Force Institute of Technology.

Lynch, W. (1991). *A Meta-analysis of pilot selection tests: Success and performance in pilot training.* Wright-Patterson Air Force Base, OH: Air Force Institute of Technology.

Melton, R. S.(Ed) (1947). *Army Air Forces aviation psychology research reports: Apparatus tests.* Washington D.C.: U.S. Government Printing Office.

Olea, M. M. & Rea, M. J. (1993). *Predicting aircrew training performance with psychometric g.* Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Pindyck, R. S. & Rubinfeld, D. L. (1991). Econometric models & economic forcast (3rd ed.). New York: McGraw-Hill.

Stead, G. (1991). "A validation study of the QANTAS pilot selection process." In E. Farmer(Ed), *Human resource management in aviation*. Aldershot, England: Avebury Technical.

Stratton, G. M., McComas, H. C., Coover, J. E., & Bagby, E. (1920). "Psychology tests for selecting aviators." *Journal of Experimental Psychology*, pp.428-433.

Street, D. R., & Dolgin D. L. (1992). *The efficacy of biographical inventory data in predicting early attrition in naval aviation officer candidate training*. Pensacola, FL: Naval Aerospace Medical Research Laboratory.

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

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center 8725 John J. Kingman Rd., Ste 0944 Ft. Belvoir, VA 22060-6218	2
2.	Dudley Knox Library Naval Postgraduate School 411 Dyer Rd. Monterey, CA 93943-5101	2
3.	Nimitz Library U.S. Naval Academy 589 McNair Rd. Annapolis, MD 21402-5029	1
4.	Superintendent United States Naval Academy Annapolis, MD 21402	1
5.	United States Naval Academy Office of Institutional Research Stop 2B Annapolis, MD 21402	1
6.	Professor Alice Crawford Naval Postgraduate School (Code SM/Cr) Monterey, CA 93943-5000	1
7.	Professor Greg Hildebrandt Naval Postgraduate School (Code SM/Hi) Monterey, CA 93943-5000	4
8.	Professor Roger Little United States Naval Academy Annapolis, MD 21402	1
9.	LT Paul Reinhart P.O. Box 1662 Annapolis, MD 21401-1662	2