AFRL-RH-WP-TR-2013-0081

CONSISTENCY OF PILOT TRAINEE COGNITIVE ABILITY, PERSONALITY, AND TRAINING PERFORMANCE IN UNDERGRADUATE PILOT TRAINING

Mark S. Teachout
University of the Incarnate Word
San Antonio, TX

Malcolm James Ree   Erica L. Barto
Operational Technologies, Corporation
San Antonio, TX

Thomas R. Carretta
Air Force Research Laboratory
Wright-Patterson AFB, OH

Raymond E. King
Headquarters, Air Force Safety Center
Kirtland AFB, NM

Charles F. Michaels
Air Education and Training Command
Randolph AFB, TX

September 2013
Interim Report

AIR FORCE RESEARCH LABORATORY
711 HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2013-0081 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//
Thomas R. Carretta
Engineering Research Psychologist
Systems Control Interface Branch
Decision Making Division

//signed//
Gregory J. Barbato
Chief, Systems Control Interface Branch
Decision Making Division

//signed//
William E. Russell
Acting Chief, Decision Making Division
Human Effectiveness Directorate
711 Human Performance Wing

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*
09-09-2013

2. REPORT TYPE
Interim

3. DATES COVERED *(From - To)*
03 June 2013 – 30 August 2013

4. TITLE AND SUBTITLE
Consistency of Pilot Trainee Cognitive Ability, Personality, and Training Performance in Undergraduate Pilot Training

5a. CONTRACT NUMBER
In-House

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Mark S. Teachout[1], Malcolm James Ree[2], Erica L. Barto[2], Thomas R. Carretta[3], Raymond E. King[4], and Charles F. Michaels[5]

5d. PROJECT NUMBER
5329

5e. TASK NUMBER
09

5f. WORK UNIT NUMBER
53290902

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of the Incarnate Word[1]
San Antonio, TX 78209

8. PERFORMING ORGANIZATION REPORT NUMBER
NA

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Material Command[3]
Air Force Research Laboratory
711 Human Performance Wing
Human Effectiveness Directorate
Warfighter Interfaces Division
Supervisory Control and Cognition Branch
Wright-Patterson AFB OH 45433-7511

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
711 HPW/RHCI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RH-WP-TR-2013-0081

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
88 ABW Cleared 10/10/2013; 88ABW-2013-4343. Report contains color.

14. ABSTRACT
Data regarding pilot trainees' cognitive ability, personality, and training outcomes collected between 1995 and 2008 from four US Air Force pilot training bases across three training tracks were examined to determine consistency in pilot aptitude and training outcomes. A sample of 9,641 pilot trainees was administered the Multidimensional Aptitude Battery (MAB) and the NEO PI–R prior to beginning Specialized Undergraduate Pilot Training. The predictors were the MAB Verbal, Performance, and Full-Scale IQ scores and the five NEO PI-R domain scores. The criterion was the C-Score percentile rank which is a weighted average of daily and check flight grades. Analyses were conducted by training track, base, and year and included descriptive statistics, mean score comparisons, and correlational analyses to determine how well the MAB and NEO PI–R predicted C-Score percentile rank. Results indicated few (continued on next page)

15. SUBJECT TERMS
Pilot trainee quality, training performance

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Antonio Ayala
Unclassified	Unclassified	Unclassified	SAR	56	**19b. TELEPHONE NUMBER *(Include Area Code)***

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

7. PERFORMING ORGANIZATIONS NAMES AND ADDRESSES (cont.):

Mark. S. Teachout[1]
University of the Incarnate Word
San Antonio, TX

Malcolm James Ree[2]   Erica L. Barto[2]
Operational Technologies, Corporation
San Antonio, TX

Thomas R. Carretta[3]
Air Force Research Laboratory
Wright-Patterson AFB, OH

Raymond E. King[4]
Headquarters, Air Force Safety Center
Kirtland AFB, NM

Charles F. Michaels[5]
Air Education and Training Command
Randolph AFB, TX

14. ABSTRACT (cont.): differences across training tracks, bases, and years and that none were substantial enough to warrant concern. Correlational results were consistent with previous research indicating that cognitive ability tests are the best predictors of performance in pilot training (Carretta & Ree, 2003; Ree & Carretta, 1996) and a variety of other occupations (Schmidt & Hunter, 1998). Overall, this research illustrated the consistency of the quality of pilot trainees as assessed by cognitive ability and personality measures, and the consistency of these measures in predicting pilot training performance over time. This overall consistency results in a more stable training system, enabling greater efficiency and effectiveness. Recommendations for future research are presented.

This page internationally left blank

Table of Contents

LIST OF TABLES

Table **Page**

PREFACE

This report describes activities performed during the examination of historical data regarding measures of US Air Force pilot candidate aptitude and training outcomes (711 HPW/RHCI), Work Unit 53290902. The opinions expressed are those of the authors and not necessarily those of the United States Government, Department of Defense, or the United States Air Force. Address all correspondence to Dr. Mark Teachout at teachout@uiwtx.edu.

1.0 SUMMARY

Data regarding pilot trainees' cognitive ability, personality, and training outcomes collected between 1995 and 2008 from four US Air Force pilot training bases across three training tracks were examined to determine consistency in pilot aptitude and training outcomes. A sample of 9,641 pilot trainees was administered the Multidimensional Aptitude Battery (MAB) and the NEO PI–R prior to beginning Specialized Undergraduate Pilot Training. The predictors were the MAB Verbal, Performance, and Full-Scale IQ scores and the five NEO PI-R domain scores. The criterion was the C-Score percentile rank which is a weighted average of daily and check flight grades. Analyses were conducted by training track, base, and year and included descriptive statistics, mean score comparisons, and correlational analyses to determine how well the MAB and NEO PI–R predicted C-Score percentile rank. Results indicated few differences across training tracks, bases, and years and that none were substantial enough to warrant concern. Correlational results were consistent with previous research indicating that cognitive ability tests are the best predictors of performance in pilot training (Carretta & Ree, 2003; Ree & Carretta, 1996) and a variety of other occupations (Schmidt & Hunter, 1998). Overall, this research illustrated the consistency of the quality of pilot trainees as assessed by cognitive ability and personality measures, and the consistency of these measures in predicting pilot training performance over time. This overall consistency results in a more stable training system, enabling greater efficiency and effectiveness. Recommendations for future research are presented.

2.0 INTRODUCTION

The selection and training of US Air Force (USAF) pilots is paramount to the success of the pilots and the USAF mission. The selection of pilots for the USAF is a vital and critical task. Not only are pilots highly valued by the USAF, they are expensive to train (Duke & Ree, 1996). The dollar costs of training are high and the risk to life and property are great. Therefore, it is important to ensure that the quality of pilot candidates remains high and stable over time, permitting pilot training to be as efficient and effective as possible. This paper examines the consistency of cognitive ability and personality measures of pilot trainees across training tracks, bases, and time.

2.1 Background

The training of USAF pilots takes place in phases and at several different locations. Some of these locations also train pilots for other military services, both US and international. For example, US Navy aviators and European or other international military train at USAF facilities.

While each location follows roughly the same training syllabus to ensure coverage of common knowledge, skills, and abilities required for success, there are differences, with the Euro-NATO Joint Jet Pilot Training (ENJJPT) program at Sheppard Air Force Base being the most divergent (King & Lochridge, 1991). The ENJJPT program is focused on training of combat pilots. Unlike Specialized Undergraduate Pilot Training (SUPT) which is taught at Columbus, Laughlin, and Vance Air Force Bases, ENJJPT has no airlift/tanker advanced training track. Also, ENJJPT students receive more hands-on flying hours in both the Primary and Advanced T-38 phases than those attending SUPT (see http://www.baseops.net/militarypilot/).

The purpose of this study was to examine the consistency of pilot trainee quality and training performance across three training tracks, four training bases, and over a 14-year time period. Maintaining a consistent high level of pilot trainee quality and training performance over time is crucial to ensuring the stability and effectiveness of the Air Force. Consistency should mean fewer changes and costs due to changes. Pilot trainee quality was measured using standardized tests of cognitive ability and personality. Training performance was measured using a composite of flying grades developed by USAF Air Education and Training Command (AETC).

2.2 USAF Pilot Candidate Selection Methods

All USAF pilot training applicants must pass the rigorous Class I flight physical standards (United States Air Force, 2011) to be eligible for selection. Medically qualified applicants are evaluated for training suitability on measures of officership and aptitude (Weeks & Zelenski, 1998). USAF Academy cadets are evaluated by Academy faculty and staff who consider academic, military, and physical performance. Applicants commissioned through the Reserve Officer Training Corps (ROTC) or Officer Training School (OTS) are administered the Air Force Officer Qualifying Test (AFOQT; Drasgow, Nye, Carretta, & Ree, 2010) and Test of Basic Aviation Skills (TBAS; Carretta, 2005). A measure of pilot training aptitude, the Pilot Candidate Selection Method (PCSM; Carretta, 2011) score, is created by combining the AFOQT Pilot composite, several TBAS subtest scores, and flying experience in a regression-weighted equation. For ROTC, medically qualified pilot training applicants are ranked on an Order of Merit score based on the PCSM score, field training, physical fitness, college grade point average (GPA), and commander's ranking. OTS pilot training candidate selection uses the "whole person" concept, where applicants receive points in three areas: experience/leadership, education/aptitude, and potential/adaptability. Regardless of commissioning source, a common theme in pilot trainee selection procedures is high intelligence, whether it involves acceptance into the USAF Academy, a high GPA, a high AFOQT score, or the impression a candidate makes on a selection board.

2.3 Medical Flight Screening

In addition to the pilot trainee selection procedures described above, all candidates must complete Medical Flight Screening (MFS; King & Flynn, 1995). The USAF MFS program screens pilot candidates prior to Specialized Undergraduate Pilot Training (SUPT). MFS includes ophthalmic and cardiac diagnostic procedures as well as several psychological tests (King, Barto, Ree, Teachout, 2011; King, Barto, Ree, Teachout, & Retzlaff, 2011), including measures of cognitive ability (MAB and MicroCog) and personality (NEO PI-R and MMPI-2-RF) tests.

2.3.1 Cognitive Tests. The primary purpose of the cognitive tests is to archive cognitive functioning data for future use in ideographic assessments where an individual is compared to themselves rather than to a collection of norms from a large population. The objective is to develop an individual registry against which future testing might be compared. Test results are particularly important for pilots seeking a waiver for return to flying status following an illness or injury that may have resulted in cognitive impairment (Chappelle, Ree, Barto, Teachout, & Thompson, 2010). During an evaluation, performance on the cognitive tests is compared with baseline scores collected prior to pilot training to determine whether any changes have occurred. Individualized (pre/post) comparisons result in more reliable return to flying duties decisions as pilots typically are very high cognitive functioning, especially in comparison to general population norms, and may remain so even after an injury or neurological event (King, 2012).

In addition to their clinical use, a recent study demonstrated that scores from the MAB and MicroCog were useful in predicting performance on several pilot training performance criteria including graduation/elimination from initial jet training and course grades (King, Carretta, Retzlaff, Barto, Ree, & Teachout, 2013). These results were consistent with prior studies of the relations of cognitive ability to pilot training performance (Carretta & Ree, 1996; Ree & Carretta, 1996).

2.3.2 Personality Tests. The Air Force does not use measures of personality for pilot training selection, but does use measures of personality based on the Big Five model (Tupes & Christal, 1961) in subsequent psychological assessments of pilots, as needed. The operational assessment tool is the Revised NEO Personality Inventory (NEO PI–R; Costa & McCrae, 1985*)*, a Big Five measure which provides domain scores on Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness.

In one of the earliest reported studies of the use of personality tests for flying personnel, Sells (1955) showed the utility of the personality constructs of “motivation to fly” and “expression of anxieties about flying.” Siem (1992) demonstrated the predictive validity of the personality constructs of hostility (r = -.12), self-confidence (r = .13), and values flexibility (r = .12) versus training completion in a sample of 509 USAF student pilots. Training graduates scored higher on self-confidence and values flexibility and lower on hostility than did those who failed due to flying training deficiency.

Anesgart and Callister (2001) examined the relationships between the NEO PI-R Big Five domain scores and success in flying training in a high-wing, propeller driven monoplane. They reported that Neuroticism, Extraversion, and Openness were related to self-elimination from the program. Boyd, Patterson, and Thompson (2005) reported statistically significant differences between the scores of pilots assigned to fly airlift/tankers versus those assigned to fly fighters for the NEO PI-R domains of Agreeableness and Conscientiousness. Fighter pilots had lower levels of Agreeableness and higher levels of Conscientiousness.

Meta-analyses (Campbell, Castaneda, & Pulos, 2010; Hunter & Burke, 1994; Martinussen, 1996) have reported modest correlations between measures of personality and pilot raining performance. Hunter and Burke (1994) reported a small correlation ($r = .10$) for personality as a predictor of flying training criteria. Martinussen (1996) reported a small correlation ($r = .14$) for personality with training completion (pass/fail). More recently, Campbell et al. (2010) performed a meta-analysis on 26 studies examining the effects of personality as a predictor of pilot training completion (pass/fail). Two higher-order personality domains, Neuroticism ($r = -.15$) and Extraversion ($r = .13$), and one lower-order facet of Neuroticism, Anxiety ($r = -.11$), were found to have an impact on training success. After correction for range restriction and reliability of the predictors, the correlations were -.25 for Neuroticism, .17 for Extraversion, and -.14 for Anxiety. The authors concluded that emotionally stable, extroverted individuals would be better able to undergo the stress of aviation training.

Finally, Chidester, Helmreich, Gregorich, and Geis, (1991) examined the relations between personality and crew coordination training performance in two samples of military pilots. Three profiles were identified through cluster analysis of the personality scales Positive Instrumental/Expressive, Negative Instrumental, and Low Motivation. These clusters replicated across samples and predicted attitude change following crew coordination training.

2.4 Purpose

The purpose of this study was to examine the consistency of pilot trainee cognitive ability, personality and training success across three training tracks, four training bases and over a 14-year time period. Maintaining a consistently high level of pilot trainee quality and training performance over time is crucial to ensuring an effective operational pilot cadre. The measures of cognitive ability were the MAB Verbal, Performance, and Full-Scale IQ scores. The measures of personality were the five NEO PI-R domain scores. The criterion was the C-Score percentile rank which is a weighted average of daily and check flight grades. Because consistency is vital to training success, fewer statistical differences are evidence of greater consistency and stability of the training system. To begin, we examined whether there were mean score differences in the MAB, NEO PI-R, and C-Score scores across the training tracks, bases, and time period. Further, we examined the predictive validity of the MAB and NEO PI-R scores for pilot training performance as measured by the C-Score. Here, consistency of prediction across tracks, bases, and time is important, as well as consistency with previous studies relating cognitive ability and personality to pilot training performance.

3.0 METHODS

3.1 Participants

A sample of 9,641 individuals selected for pilot training was administered the MAB and the NEO PI–R prior to beginning the 53 week Specialized Undergraduate Pilot Training (SUPT) program. All participants were college graduates or were near completion of college at time of testing. Selection ratios for pilot training assignments vary from year to year as a function of the number of applicants and the number of training positions available for each commissioning source. Of the participants reporting demographics information, all were under the age of 36 years, with a modal age of 22 years, mean age of 24 years, and standard deviation of 2.6 years. Most of the participants (93%) were men. Racial and ethnic distributions indicated that 91% were white, 2% were African American, 3% were Hispanic, and 4% were "other." All were tested at either the School of Aerospace Medicine at Brooks City-Base, TX or at the United States Air Force Academy in Colorado Springs, CO.

3.2 Measures

3.2.1 Multidimensional Aptitude Battery (MAB). The MAB (Jackson, 1984) is a broad-based test of intellectual ability patterned after the Wechsler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981). The MAB has 10 subtests that are combined to produce three summary scores: verbal IQ (VIQ), performance IQ (PIQ), and full-scale IQ (FSIQ). Previous research has demonstrated that the full-scale IQ score for the MAB and WAIS-R are strongly correlated (r = .91; Conoley & Kramer, 1989; Jackson, 1984) and that the MAB measures general mental ability in several age groups (Carretta, Retzlaff, Callister, & King, 1998; Kranzler, 1991; Lee, Wallbrown, & Blaha, 1990; Wallbrown, Carmin, & Barnett, 1988, 1989). The MAB requires less than 1.5 hours to administer and can be individually or group administered. The subtests each have a normative mean of 50 and standard deviation of 10. FSIQ, VIQ, and PIQ scores have a mean of 100 and a SD of 15 in the general population. MAB norms are based on a sampling of 9 age groups that were diverse in terms of sex, ethnicity, race,

Table 1. MAB-II Subtest and Summary Score Descriptions

Scale	Subtest	Description	Reliability
VIQ, FSIQ	Information	Assesses the extent to which an individual has acquired knowledge about diverse topics	.87
VIQ, FSIQ	Comprehension	Measures the ability to evaluate social behavior, identify behavior that is more socially acceptable, and provide reasons why certain social customs and laws are practiced	.88
VIQ, FSIQ	Arithmetic	Assesses reasoning and problem solving ability through the solution of numerical problems	.80
VIQ, FSIQ	Similarities	Assesses the ability to conceptualize properties of an object and to compare them to those of another object, identifying the most similar characteristic	.90
VIQ, FSIQ	Vocabulary	Measures the ability to identify word meaning	.88
PIQ, FSIQ	Digit Symbol	Assesses visual-motor activity in substituting symbols for digits	.95
PIQ, FSIQ	Picture Completion	Measures the ability to identify missing elements in a picture	.88
PIQ, FSIQ	Spatial	Assesses the ability to visualize abstract objects in different positions in two-dimensional space	.96
PIQ, FSIQ	Picture Arrangement	Assesses the ability to arrange a set of randomly ordered pictures into a meaningful sequence	.85
PIQ, FSIQ	Object Assembly	Measures the ability to identify a complete object from disassembled	.89

Note. Reliability was estimated through internal consistency using KR-20 (Jackson, 2003).

and North American (Canada and United States) geographic location. Test-retest reliability for the IQ scores ranges from .94 to .98 (Jackson, 1998) for an average retest interval of 45 days.

Table 1 provides brief descriptions of the subtests and indicates the summary IQ scores to which they contribute. Internal consistency reliability of the MAB–II in a sample of 91 twenty year olds was estimated using KR-20 (Jackson, 2003). This age group was the most similar to our participants. Reliabilities of the IQ scores ranged from .97 to .98 and reliabilities of the subtests ranged from .80 to .96.

3.2.2 NEO PI–R. The NEO PI-R (Costa & McCrae, 1985) was designed to measure the Big Five personality domains and the facets or traits that underlie each domain. The five domains are Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each domain consists of six subscales called facet scores. These domains and facets provide a comprehensive measurement of adult personality.

Table 2. NEO-PI-R Domain Definitions and Reliabilities

Test	Definition	Reliability
Neuroticism (N)	The tendency to experience negative emotions (anger, sadness, fear) and be emotionally unstable	.92
Extraversion (E)	The enjoyment of social situations, excitement, and stimulation	.89
Openness to Experience (O)	A willingness to explore new ideas and values; desire for aesthetics	.87
Agreeableness (A)	The desire to sympathize with and help others	.86
Conscientiousness (C)	Seeking a high-level of organization and planning; the tendency to plan carefully and exercise self-discipline	.90

Note. Reliability was estimated through internal consistency using Coefficient alpha for a developmental sample of 1,539 respondents (Costa & McCrae, 1985).

The NEO PI–R was developed with the goal of being a multipurpose personality inventory useful for predicting many criteria such as behaviors related to illness, career interests, psychological health, and styles of coping (Costa & McCrae, 1985). It contains 240 statements that require examinees to respond on a Likert-type scale, ranging from 1 - “strongly disagree” to 5 - “strongly agree.” Table 2 shows a description of the five domain scales as well as their internal consistency reliabilities in a sample of 1,539 men and women in a large organization. Reliability coefficients for the 30 facets are reported in the test manual and range from .56 to .81 (Costa & McCrae, 1985). For the current study, the normative sample for adults served as the normative reference and the test was administered and scored via computer (Costa & McCrae, 1985).

3.2.3 Training Performance Criterion. Specialized Undergraduate Pilot Training (SUPT) consists of a Primary Aircraft Training phase and an Advanced Aircraft Training phase. Primary Aircraft Training (T-6) consists of about 90 hours of flight training instruction over 22 weeks. The purpose is to teach basic flying skills including contact, instruments, formation (2-ship), and navigation. At the end of this phase, students are assigned to advanced training in either the fighter/bomber or the airlift/tanker track. Advanced training track assignments are a function of student preferences, training performance, instructor ratings, and aircraft availability. The fighter/bomber advanced training track (T-38) includes about 120 hours of flight instruction over 24 weeks designed to prepare students for follow-on fighter/bomber training assignments. The initial training focus is on contact, instruments, formation (2/4 ship), navigation, and low level flight. The airlift/tanker advanced training track (T-1) has about 115 hours of flight instruction over 26 weeks. The purpose is to prepare students for assignments to multiengine jet and turboprop aircraft. The training focuses on transition, instruments, navigation, low level, and formation. It should be noted that training at Sheppard AFB differs from that at the other three bases. Sheppard AFB hosts the Euro-NATO Joint Jet Pilot Training (ENJJPT) program which is focused on training of combat pilots. It has no airlift/tanker advanced training track. Also, ENJJPT students receive more flying hours in both the Primary (125 hours over 26 weeks) and Advanced T-38 (135 hours over 26 weeks) than those attending SUPT.

The C-Score is a standardized flying training performance criterion measure developed by Headquarters Air Education and Training Command (AETC) in order to provide compatibility and comparability of performance at all US Air Force pilot training bases. The C-Score was developed after it was determined that there were mean differences in the ratings and other measures of pilot training performance across bases. For example, a very high scoring pilot at Base A might be scored lower than a high scoring pilot at Base B, due to idiosyncratic rating behavior by an instructor and/or check ride raters. As a result, comparisons across bases from one pilot training class to another were uncertain.

To enable meaningful comparisons (base-to-base, class-to-class, year-to-year, and pilot-to-pilot), the C-Score is a percentile rank based on a two year moving average. This allows the C-Score to reflect the training performance of each pilot, relative to the previous two years of training performance for all pilots. Using past pilot performance as a moving-baseline average produces more reliable, stable, and interpretable scores, while permitting distinctions between individual performance.

The C-Score uses daily flying grades and check flight grades weighted approximately 1 to 2 in favor of the check flights. Daily flying grades include instructor pilots' evaluations of a pilot trainee's performance on all flights other than check flights. Daily flying grades are a weighted average of all flying training procedures/maneuvers performed during a flight and are rated unsatisfactory, fair, good, and excellent. In addition to daily flights, during training, pilot trainees must pass a check flight for each course of instruction. As with daily flying grades, check flight grades are a weighted average of ratings of flying procedures/maneuvers, which may have values of unsatisfactory, fair, good, and excellent. Maneuver grade point values are weighted based on the importance of the maneuver.

The C-Score calculation is standardized against approximately 200 previous students at that particular base or two years of students, whichever is greater. The calculations for each class are based on a "moving average," as one class is added to the population the oldest class is eliminated from the population. The C-Score is calculated for each class. Students are ranked on their C-Score value and each student is given a C-Score percentile rank, a number between 0% and 100%. The C-Score and percentile rank for a student are only recorded when the student is part of the graduating class.

3.3 Procedures

Student pilots' scores from the MAB and NEO PI–R were matched to their C-Scores, computed based on ratings received during SUPT. Descriptive statistics were computed for those in each training track (Primary, Advanced T-38, and Advanced T-1), permitting statistical comparisons of mean differences. MAB IQ scores and NEO PI–R domain scores were correlated with C-Score percentile ranks by training track for each aircraft flown during training, by the USAF base conducting the training, and for each year of training.

Because the sample participants were a highly selected group, all statistics computed on their scores suffer from range restriction. Restriction of range occurs when samples are "restricted" on one or more variables. This result occurs when previous selection of the "best" people reduces the usual range of scores on any variables measured. This causes statistical estimates to be biased with respect to population values. Corrections for range restriction have been available for 100 years (Pearson, 1903) and 70 years for the multivariate case (Lawley, 1943). The assumptions underlying range restriction correction are the same as two of the three assumptions underlying the computation of a Pearson product-moment correlation. These two assumptions are linearity of form and homoscedasticity. If the assumptions are met to estimate the correlation coefficient, the assumptions are met to compute the correction. Restriction of range generally causes statistical indexes to underestimate true values.

Using the method of Lawley (1943), the multivariate correction of the previously computed correlations was conducted. The normative sample of the MAB–II and NEO PI–R provided the means, standard deviations, and correlations used for the correction. The corrected means, standard deviations, and correlations are superior estimates of the population values compared to the uncorrected values. This method removes the bias from the uncorrected sample estimates.

3.4 Analyses

Analyses were conducted by training track, base, and year. Three analyses were conducted for each of these sets. First, descriptive statistics were calculated for the MAB IQ Scores, NEO PI–R domain scores, and C-Score percentile rank. Second, analyses (*t*-tests or one-way ANOVAs) were conducted to determine statistical differences in mean scores for each variable in each category. Third, correlational analyses were conducted to determine how well the MAB and NEO PI–R scores predicted C-Score percentile rank. Each correlational analysis displays the uncorrected correlations and the correlations corrected for range restriction. Sample sizes differ for each analysis and are noted below each table. All analyses used a one-tailed test. The analyses that involved year-to-year comparisons used a .01 Type I error rate due to the large number of comparisons. All other analyses used a .05 Type I error rate. It should be noted, that while the very large samples used in this study ensure sufficient statistical power, very small differences will be statistically significant yet may offer little practical predictive power. Importantly, fewer statistical differences across training tracks, bases, and years are desirable, as this indicates greater stability and consistency in the measures.

4.0 RESULTS

4.1 Analyses by Training Track

The first set of analyses was conducted by training track, Primary, Advanced T–38, and Advanced T–1. Data were collapsed across training bases and years for these analyses.

4.1.1 Means. Descriptive statistics are shown in Table 3. The MAB IQ scores for the student pilots were severely range restricted compared to the normative values where the means and SDs are 100 and 15. The IQs for each of the training groups were high at about 120 (about 1.33 SD above the normative mean) and the variances of the scores were much less than the normative values. For the FSIQ score the variance for the trainees was about 18% of the normative value. The results were mixed for the NEO PI-R where trainees were above the normative mean score of 50 for Extraversion and Conscientiousness and below the normative mean for Neuroticism and Agreeableness.

Independent-groups *t*-tests were conducted on each of the nine variables to identify significant differences between the two advanced training tracks. Because the advanced tracks include the students from the primary track, no comparisons were made with the primary track. Results indicated that there were small, but statistically significant mean differences between the T-38 and T-1 advanced tracks for 4 of the 9 scores. Cohen (1988) characterizes standardized mean differences (*d*) of .2 as small, .5 as medium, and .8 or greater as large. All mean score differences between trainees in the T-38 and T-1 tracks were small. T-38 trainees scored higher on the MAB VIQ (d = .31) and FSIQ (d = .20) scores than did T-1 trainees. However, T-38 trainees scored lower on the NEO PI-R Agreeableness score (d = -0.16) and the C-Score (d = -0.17) than those in the T–1 track.

4.1.2 Correlations. Table 4 summarizes the correlational analyses by training track. Cohen (1988) characterizes correlations of .10 as small, .30 as medium, and .50 or greater as large. All of the observed correlations between the MAB-II and NEO PI-R with the C-Score criterion were small. Even after correction for range restriction none of the correlations with the C-Score exceeded .30. Overall, the magnitudes of the correlations were higher for the MAB compared to the NEO PI–R. All of the MAB IQ correlations with C-Score percentile rank were statistically significant for each training phase. Eight of the 15 correlations between the NEO PI-R scores and C-Score were statistically significant.

Table 3. Descriptive Statistics for Primary and Advanced Training Tracks

Score	Primary		Advanced T-38		Advanced T-1		T-38 vs. T-1	
	Mean	SD	Mean	SD	Mean	SD	*d*	*t*
C-Score	0.52	0.29	0.49	0.29	0.54	0.29	-0.17	-5.56**
VIQ	119.03	6.57	120.18	6.31	118.19	6.36	0.31	10.15**
PIQ	119.41	8.17	120.62	7.90	120.27	7.75	0.04	1.43
FSIQ	120.58	6.50	121.83	6.29	120.56	6.17	0.20	6.55**
N	46.65	9.37	46.06	9.46	46.29	9.17	-0.02	-0.79
E	57.59	9.56	58.12	9.66	57.65	9.47	0.05	1.58
O	50.67	10.18	50.49	10.39	50.05	9.66	0.04	1.39
A	43.81	10.56	42.70	10.71	44.38	10.28	-0.15	-5.12**
C	54.73	10.17	55.50	10.04	55.60	9.86	-0.01	-0.32

Note: Primary N = 9,396, Advanced T-38 N = 3,297, Advanced T-1 N = 1,524
*$p < .05$; **$p < .001$

Table 4. Corrected and Uncorrected Correlations of MAB-II IQ Scores and NEO-PI-R Domain Scores with C-Score Percentile Rank by Training Track

	Primary		Advanced T-38		Advanced T-1	
Score	U	C	U	C	U	C
VIQ	.092*	.240	.105*	.270	.103*	.250
PIQ	.120*	.260	.111*	.270	.056*	.190
FSIQ	.128*	.285	.127*	.292	.092*	.219
N	-.023*	-.040	.014	-.020	.020	-.140
E	.008	-.060	.038*	-.050	-.002	-.090
O	-.064*	.050	-.067*	.070	-.042	.060
A	-.019	-.030	-.059*	-.060	.029	.030
C	.031*	.000	.043*	.020	.107*	.070

Note. Sample sizes were Primary N = 9,396, Advanced T-38 N = 3,297, Advanced T-1 N = 1,524. U means uncorrected and C means corrected for range restriction. The multivariate correction method (Lawley, 1943) was used for all scores except the MAB FSIQ which used the univariate (Thorndike, 1949) method. FSIQ is a linear combination of the VIQ and PIQ scores and the corrected correlations cannot be estimated using the multivariate (Lawley, 1943) method.
*$p < .05$

4.2 Analyses by Base

The second set of analyses was conducted by base (Columbus, Laughlin, Sheppard,[1] and Vance), for Primary, Advanced T–38, and Advanced T–1 training. The sample sizes for the training tracks for each base are provided in Table 5. Due to space limitations, the tables summarizing these analyses cannot be presented here.

Table 5. Sample Sizes for Training Track by Base

	Training Track		
Base	Primary	Advanced T-38	Advanced T-1
Columbus	2,781	780	589
Laughlin	2.911	840	584
Sheppard	1,024	1,006	NA
Vance	2,320	650	351

[1] Sheppard AFB, which hosts the combat-oriented Euro-NATO Joint Jet Pilot Training (ENJJPT) program, does not have an Advanced T-1 training track.

4.2.1 Means. Descriptive statistics were calculated for each of the 9 variables. One–way analyses of variance were conducted to identify any statistically significant differences among the bases for Primary, T-38, and T-1 training.

4.2.1.1 Primary training. Results indicated that there were small (Cohen, 1988), but statistically significant mean score differences between bases for six variables. The mean C-Score percentile rank for Laughlin AFB (M = 0.54) was significantly higher than for Columbus (M= 0.50) and Vance (M = 0.51) which were equivalent to small standardized differences (d = 0.13 and 0.10, respectively). Sheppard AFB differed from the other bases on five scores. Trainees at Sheppard were about 2 points higher on all three MAB IQ scores than trainees at the other bases. For example, the standardized mean difference (d) on the FSIQ score ranged from 0.33 to 0.39. Further, trainees at Sheppard were significantly lower on Agreeableness (about 1 point or 0.10 d) and higher on Conscientiousness (about 3 points or 0.31 d) than trainees at the other bases.

4.2.1.2 Advanced T-38 training. There were small, but statistically significant mean differences among the bases for 4 of the 9 variables. Where significant differences occurred, they were between Sheppard and one or more of the other bases. When differences occurred for the MAB the scores at Sheppard were higher than for the other bases. The mean VIQ score at Sheppard was higher than for Laughlin and Vance. The mean PIQ score was higher than Laughlin and the mean FSIQ score was higher than each of the other three bases. The mean score on Conscientiousness at Sheppard was significantly higher than the other three bases.

4.2.1.3 Advanced T-1 training. Results indicated that there were small, but statistically significant mean differences among bases for only the C-Score percentile rank mean. It was significantly lower for Columbus (M = 0.47), compared to Laughlin (M = 0.57, d = -0.28) and Vance (M = 0.58, d = -0.30).

4.2.2 Correlations. The pattern of correlations between the MAB and NEO PI-R scores and C-Score by Base was similar to those observed when the data were collapsed across bases (see Table 4).

4.2.2.1 Primary training. For each base, all three MAB IQ scores demonstrated small, but statistically significant relations to the C-Score. For example, the observed correlations between the MAB FSIQ and C-Score ranged from .131 (Columbus) to .137 (Sheppard) and from .290 (Columbus) to .327 (Sheppard) after correction for range restriction. The relations between the NEO PI-R scores and the C-Score were weaker than those for the MAB. Only 7 of 20 correlations were statistically significant and did not produce a consistent pattern across bases.

4.2.2.2 Advanced T-38 training. Results for T-38 training were less consistent than those for Primary training. None of the MAB scores were significantly correlated with performance in T-38 advanced training at Columbus AFB. The observed correlation between the MAB FSIQ and C-Scores ranged from .058 (Columbus) to .169 (Laughlin) and from .130 (Columbus) to .350 (Laughlin) after correction for range restriction. As with Primary training, the correlations between the NEO PI-R scores and C-Score were weaker than those for the MAB. Only 7 of 20 NEO PI-R/C-Score correlations were statistically significant. Three of the 7 correlations were for Openness with values that ranged from -.009 (ns) (Laughlin) to .113 (Vance).

4.2.2.3 Advanced T-1 training. As with T-38 training, results for T-1 training were less consistent than those for Primary training. None of the MAB scores were significantly related to performance in T-1 advanced training at Columbus AFB. The observed correlation between the MAB FSIQ and C-Scores ranged from .056 (Columbus) to .135 (Vance) and from .134 (Columbus) to .310 (Vance) after correction for range restriction. The MAB PIQ score was not related to training performance for T-1 training at any of the bases.

Overall, the magnitude of the correlations was higher for the MAB compared to the NEO PI–R. Only 2 of the 15 observed correlations between the NEO PI-R scores and the C-Score were statistically significant. Both were for Conscientiousness at Laughlin (r = .091) and Vance (r = .179). After correction for range restriction the correlations were .040 (Laughlin) and .220 (Vance).

4.3 Analyses by Year

The third set of analyses was conducted by year (1995–2008) for each training phase. Due to the large number of tables needed to summarize these analyses, they are presented in Appendix A.

4.3.1 Means. A one–way analysis of variance was conducted on each of the 9 scores to determine statistically significant differences among the 14 years for each training phase. The numerous comparisons for these analyses (91 comparisons for each of 9 scores for each phase = 819 comparisons/phase) should be viewed with caution, due to the increased likelihood of Type I error, that is, finding significant differences by chance as the number of comparisons increases. For this reason, a $p < .01$ level of significance was used for comparing these mean differences. Further, rather than reporting and interpreting all of the significant differences, we focused on data trends.

4.3.1.1 Primary training. Results indicated that there were statistically significant differences for 8 of the 9 scores for Primary training. Overall, while there were some statistically significant differences (75/819 = 9.1%), the scores were very stable, indicating that the characteristics and quality of pilot trainees were consistent over time. The number of significant differences was

Table 6. Number of Significant Mean Score Differences Across Years

Score	Training Phase		
	Primary	Advanced T-38	Advanced T-1
C-Score	13	11	3
VIQ	3	0	0
PIQ	22	7	0
FSIQ	5	0	0
Neuroticism	9	1	0
Extraversion	0	2	0
Openness	2	0	0
Agreeableness	11	3	1
Conscientiousness	10	3	0
TOTAL	75	27	4

Note. The numbers indicate the number of statistically significant mean score differences at the $p < .01$ level.

largest for the MAB PIQ score (22/91 = 24.1%) and C-Score (13/91= 14.3%). See Table 6. Sixteen of the 22 significant differences for PIQ were for years 2001-2003, where the PIQ scores were lower than for other years. For the C-Score, the mean for 1997 was higher than that for 1999 and 2003-2006 and the mean for 2002 was higher than those for 2003-2006.

4.3.1.2 Advanced T-38 training. The degree of consistency in mean scores was greater for advanced training than for Primary training. Six of the 9 scores exhibited significant differences for T-38 training. Only 3% (25/891) of the comparisons reached statistical significance. As with Primary training, most of the significant differences occurred for the C-Score (11) and MAB PIQ (7). For the C-Score, 10 of the 11 differences occurred for 2000-2001 which were lower than other years. The MAB PIQ scores for 2005-2006 were higher than those for 2000-2003.

4.3.1.3 Advanced T-1 training*[2]*. Only 2 of 9 scores showed statistically significant mean score differences across year of training. Only 7.4% of the comparisons (4/54) were statistically significant indicating a remarkable degree of consistency in scores for the T-1 trainees.

4.3.2 Correlations. The correlational results broken out by year of training were consistent with those reported earlier where the data were collapsed across years of training. Overall, the magnitude of the correlations with the C-Score were higher for the MAB IQ scores than for the NEO PI–R domain scores.

4.3.2.1 Primary training. Although there was some variability, the magnitude of the correlations between the MAB and NEO PI-R scores with the C-Score across years was consistent and mirrored the results summed across years. Overall, the magnitude of the correlations was higher for the MAB compared to the NEO PI–R.

4.3.2.2 Advanced T-38 training. Again, the results broken out by year were consistent with those accumulated across years of training. The magnitude of the MAB correlations with the C-Score were higher than those for the NEO PI-R.

4.3.2.3 Advanced T-1 training. Consistent with previous analyses, overall, the magnitude of the correlations with the C-Score was higher for the MAB than for the NEO PI–R. Further, there was little variability by year.

5.0 DISCUSSION

Consistency is vital to training success. The purpose of this study was to examine the extent to which pilot trainee quality and training performance were consistent over training track, training location, and time. Trainee quality was assessed using scores from the Multidimensional Aptitude Battery and the NEO PI-R. Trainee performance was measured using the C-Score percentile rank, a broad measure of performance based on daily and check flight grades.

5.1 Training Tracks

Results indicated there were small, but statistically significant mean differences between the students in the tanker/transport (T-1) and fighter/bomber (T-38) advanced training tracks for two of the cognitive and one of the personality scores. The mean VIQ (T-38 = 120.18, T-1 = 118.19, d = 0.31) and FSIQ (T-38 = 121.83, T-1 = 120.56, d = 0.20) scores were slightly higher for the

[2] T-1 training began in 2005. Prior to 2005 a different aircraft was used in tanker/transport training.

T–38 track. This is consistent with the selection and assignment of pilots for advanced training and with prior studies (Boyd et al., 2005). Because the T–38 track leads to more preferred assignments in fighter and bomber aircraft, it is desirable that students with higher cognitive ability tend to be assigned to this track. Although these results were statistically significant, it is not clear whether they have any practical importance. The mean cognitive ability for both tracks was very high compared to normative values and the mean score differences between those assigned to the fighter/bomber and tanker/transport tracks were small (1.27 points for the FSIQ or .20 *d*). The predictive utility of these scores for improving advanced training assignment should be evaluated in the context of other assignment factors such as Preliminary training (T-6) performance and student preferences.

The lower scores on Agreeableness for the T–38 track (T-38 = 42.70, T-1 = 44.38, d = -0.15) were consistent with previous results on personality for the highly selected pilot population. For example, pilots score lower on Agreeableness than the general population (King, Barto, Ree, & Teachout, 2011). Again, the practical utility of this finding should be evaluated in the context of advanced training assignment factors such as T-6 performance, instructor recommendations, and student preferences.

The overall correlational results for training tracks indicated that cognitive ability was related to pilot training success, and these correlations were generally higher than those for the personality measures. The predictiveness of cognitive ability for pilot training performance is consistent with the research literature (Carretta & Ree, 2003; Ree & Carretta, 1996). The relative strength of the predictiveness of cognitive and personality measures is consistent with meta-analytic studies regarding the predictiveness of commonly used selection methods (Schmidt & Hunter, 1998).

5.2 Training Bases

The most consistent result for comparisons of trainee quality across training bases was that Sheppard AFB had higher quality pilot trainees based on higher cognitive ability scores and higher scores on Conscientiousness, a key personality trait predictive of success in all jobs (Barrick & Mount, 1991). These pilots also were lower on Agreeableness. Further examination of student assignment to different bases is warranted to understand these differences. We can only speculate as to the underlying cause of these relations. Sheppard AFB is where the Euro-NATO Joint Jet Pilot Training (ENJJPT) program is located. ENJJPT is a combat-oriented program. There is no separate advanced training track for non-fighter pilots. As a result, it is likely that pilot candidates who are considered to have a high probability of becoming fighter-qualified are assigned to ENJJPT.

5.3 Year

Given the substantial number of year-to-year comparisons made, the number of statistically significant differences was extremely small. This result illustrates the consistency of pilot trainee scores on cognitive ability and personality measures over time. With pilot trainee characteristics this stable, fewer disruptions and adjustments are needed, the training system is more stable, enabling greater efficiency and effectiveness. There were more year-to-year differences noted in the C-Score measure, suggesting that more research is warranted on its stability.

6.0 CONCLUSIONS AND RECOMMENDATIONS

Overall, these results convey two notable messages. First, the quality of USAF pilot trainees has been remarkably consistent across training tracks and training locations over a 14 year time period. This is likely a function of the availability of sufficient numbers of high quality applicants to fill available training positions and consistency in selection and training methods. Second, consistent with prior studies, measures of cognitive ability and personality are important determinants of pilot training success.

Despite their proven relations with pilot training outcomes, current USAF selection and classification methods do not leverage measures of cognitive ability and personality in an optimal manner. Although cognitive ability is represented in USAF pilot trainee selection methods such as the AFOQT and PCSM, measures of personality are not. Further, neither measures of cognitive ability or personality are considered when making advanced training assignments. To this end, we suggest that studies be conducted to examine the incremental validity of personality for USAF pilot training qualification when used in combination with the PCSM score and that measures of pilot aptitude (e.g., AFOQT, PCSM, NEO PI-R) be examined to determine their utility in improving advanced training assignments when used in combination with Preliminary training performance, instructor ratings, and student preferences.

7.0 REFERENCES

Anesgart, M. N. & Callister, J. D. (2001). *Predicting training success with the NEO PI–R: The use of logistic regression to determine the odds of completing a pilot screening program,* AFRL-HE-WP-TR-2001-0074. U.S. Air Force Armstrong Laboratory, Brooks Air Force Base, TX.

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

Boyd, J. E., Patterson, J. C., & Thompson, B. T. (2005). Psychological test profile of USAF pilots before training vs. type aircraft flown. *Aviation, Space, and Environmental Medicine, 76*, 463–468.

Campbell, J. S., Castaneda, M., & Pulos, S. (2010). Meta-analysis of personality assessments a predictors of military aviation training success. *International Journal of Aviation Psychology, 20,* 92–109.

Carretta, T. R. (2005). *Development and validation of the Test of Basic Aviation Skills (TBAS),* AFRL-HE-WPTR-2005-0172. Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, OH.

Carretta, T. R. (2011). Pilot Candidate Selection Method: Still an effective predictor of US Air Force pilot training performance. *Aviation Psychology and Applied Human Factors, 1,* 3-8.

Carretta, T. R., & Ree, M. J. (1996). U. S. Air Force pilot selection tests: What is measured and what is predictive? *Aviation, Space, and Environmental Medicine, 67,* 275-283.

Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. In P.S. Tsang & M.A. Vidulich (Eds.), *Principles and practice of aviation psychology (*pp. 357-396). Mahwah, NJ: Erlbaum.

Carretta, T. R., Retzlaff, P. D., Callister, J. D., & King, R. E. (1998). A comparison of two U.S. Air Force pilot aptitude tests. *Aviation, Space, Environmental Medicine*, *69*, 931 – 935.

Chappelle, W., Ree, M. J., Barto, E. L., Teachout, M. S., & Thompson, W. T. (2010). *Joint use of the MAB-II and MicroCog for improvements in the clinical and neuropsychological screening and aeromedical waiver process of rated USAF pilots,* AFRL-SA-BR-TR-2010-0002. U.S. Air Force School of Aerospace Medicine, Brooks City-Base, TX.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Mahwah, NJ: Erlbaum.

Conoley, J. C., & Kramer, J. J. (1989). *The tenth mental measurements yearbook*. Lincoln, NE: The University of Nebraska Press.

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources.

Drasgow, F., Nye C., Carretta, T. R., & Ree, M. J. (2910). Factor structure of the Air Force Officer Qualifying Test Form S: Analysis and comparison with previous forms. *Journal of Military Psychology, 22,* 68-85.

Duke, A. P., & Ree, M. J. (1996). Better candidates fly fewer training hours: Another time testing pays off. *International Journal of Selection and Assessment, 4*, 115-121.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology*. Vol. 2 (pp. 141-165). Beverly Hills, CA: Sage.

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

Jackson, D. N. (1984). *Multidimensional Aptitude Battery – Manual.* London, OT: Research Psychologists Press.

Jackson, D. N., (1998). *Multidimensional Aptitude Battery-II: Manual.* Port Huron, MI: SIGMA Assessment Systems.

Jackson, D. N. (2003). *Multidimensional Aptitude Battery-II Manual.* Port Huron, MI: SIGMA Assessment Systems, Inc.

King, R. E. (2012). Psychological testing for mental health screening, suitability determinations, and archival purposes to improve safety and reduce costs. In A. Droog & M. Heese (Eds.), *Proceedings of the 30th Conference of the European Association for Aviation Psychology* (pp. 51-56). Sardinia, Italy.

King, R. E., Barto, E., Ree, M. J., & Teachout, M. S. (2011). *Compilation of pilot personality norms,* AFRL-SA-WP-TR-2011-0008. U.S. Air Force School of Aerospace Medicine, Aeromedical Research Department, Wright-Patterson AFB, OH.

King, R. E., Barto, E., Ree, M. J., Teachout, M. S., & Retzlaff, P. D. (2011). *Compilation of pilot cognitive ability norms,* AFRL-SA-WP-TR-2012-0001. U.S. Air Force School of Aerospace Medicine, Aeromedical Research Department, Wright-Patterson Air Force Base, OH.

King, R. E., Carretta, T. R., Retzlaff, P. D., Barto, E., Ree, M. J., & Teachout, M. S. (2013). Standard cognitive psychological tests predict military pilot training outcomes. *Aviation Psychology and Applied Human Factors, 3,* 28-38.

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

King, R. E. & Lochridge, G. K. (1991). Flight psychology at Sheppard Air Force Base. *Aviation, Space, and Environmental Medicine, 62,* 1185-1188.

Lawley, D.N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh,* LXII-Part 1, 19-32.

Lee, M. S., Wallbrown, F. H., & Blaha, J. (1990). Note on the construct validity of the Multidimensional Aptitude Battery. *Psychological Reports, 67,* 1219–1222.

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology, 6,* 1-20.

Pearson, K. (1903). Mathematical contributions to the theory of evolution: II. On the influence of natural selection on the variability and correlation of organs. *Royal Society Philosophical Transactions, 200 (Series A),* 1-66.

Ree, M. J., & Carretta, T. R. (1996). Central role of *g* in military pilot selection. *International Journal of Aviation Psychology*, *6*, 111-123.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274.

Sells, S. B. (1955). Development of a personality test battery for psychiatric screening of flying personnel. *Journal of Aviation Medicine, 26*, 35–45.

Siem, F. M. (1992). Predictive validity of an automated personality inventory for Air Forcc pilot selection. *International Journal of Aviation Psychology, 2*, 261–270.

Thorndike, R. L. (1949). *Personnel selection.* NY: Wiley.

Tupes, E. C., & Christal, R. C. (1961). *Recurrent personality factors based on trait ratings.* ASD-TR-61-97. Personnel Laboratory, Aeronautical Systems Division, Lackland AFB, TX.

United States Air Force (2011). *Medical examinations and standards,* Air Force Instruction 48-123. Washington, DC: Department of the Air Force.

Wallbrown, F. H., Carmin, C. N., & Barnett, R. W. (1988). Investigating the construct validity of the Multidimensional Aptitude Battery. *Psychological Reports, 62,* 871–878.

Wallbrown, F. H., Carmin, C. N., & Barnett, R. W. (1989). A further note on the construct validity of the Multidimensional Aptitude Battery. *Journal of Clinical Psychology, 45,* 429–433.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scales – Revised (WAIS-R).* NY: The Psychological Corporation.

APPENDIX A
Tables for Analyses by Year

Table A-1. Descriptive Statistics by Year for Primary Training

		C-score		VIQ		PIQ		FSIQ		N		E		O		A		C	
Year	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1995	17*	0.57	0.33	119.82	8.53	120.41	11.74	121.71	9.53	42.82	8.05	59.88	9.58	51.94	12.73	48.65	10.40	58.00	6.56
1996	149	0.55	0.30	117.96	6.95	119.34	9.28	120.07	7.05	45.36	9.14	57.09	9.19	50.77	9.39	41.26	10.75	55.68	9.30
1997	334	0.58	0.29	118.41	6.57	118.72	8.69	119.91	6.78	46.36	9.80	57.66	9.50	50.68	10.80	42.49	10.36	54.27	10.39
1998	699	0.52	0.28	118.97	6.61	119.00	8.24	120.33	6.55	47.42	9.49	57.02	9.51	50.66	10.48	44.11	11.02	54.93	10.27
1999	891	0.50	0.30	120.02	6.70	119.87	8.10	121.39	6.52	46.25	9.42	57.05	9.81	51.06	10.13	44.81	10.64	54.50	10.02
2000	1031	0.51	0.29	119.33	6.55	118.98	8.12	120.56	6.48	46.13	9.00	56.86	9.34	51.27	10.39	44.75	10.73	54.55	10.20
2001	1055	0.53	0.29	119.36	6.53	118.40	7.95	120.23	6.35	46.88	9.35	57.48	9.25	51.52	10.02	44.59	11.05	54.17	9.97
2002	983	0.55	0.29	118.97	6.65	117.71	8.65	119.68	6.83	48.05	9.74	57.78	9.54	51.27	10.79	43.16	10.19	52.98	10.82
2003	987	0.49	0.29	118.91	6.95	118.65	8.06	120.13	6.71	47.85	9.44	57.78	9.55	50.97	10.21	42.84	10.10	53.40	10.33
2004	457	0.46	0.29	119.01	6.50	120.42	8.49	121.08	6.77	46.63	9.19	57.94	9.79	49.57	10.15	41.69	10.42	55.07	10.04
2005	374	0.48	0.30	118.20	5.91	121.11	8.14	121.02	6.16	54.70	10.32	58.74	8.95	48.92	9.73	42.59	10.98	56.10	10.04
2006	640	0.48	0.29	118.38	6.16	120.91	7.57	120.99	5.96	46.01	8.31	58.59	9.56	50.20	9.34	43.06	9.64	56.03	9.50
2007	877	0.54	0.29	118.60	6.37	120.37	7.71	120.82	6.14	45.80	8.99	57.94	9.82	49.80	10.09	44.18	10.43	56.21	9.92
2008	902	0.54	0.29	119.18	6.48	120.63	7.70	121.26	6.28	46.05	9.51	57.40	9.81	50.00	9.69	44.95	10.47	55.82	9.97

Table A-2. Differences by Year for the C-Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997					Rc				Rc	Rc	Rc	Rc		
1998														
1999														
2000														
2001										Rc				
2002									Rc	Rc	Rc	Rc		
2003														
2004														
2005														
2006														
2007									Rc	Rc		Rc		
2008										Rc		Rc		

Note. An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-1 to compare the actual means for each year.

Table A-3. Differences by Year for the MAB VIQ Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999											Rc	Rc	Rc	
2000														
2001														
2002														
2003														
2004														
2005														
2006														
2007														
2008														

Note. An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-1 to compare the actual means for each year.

Table A-4. Differences by Year for the MAB PIQ Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999							Rc							
2000								Rc						
2001														
2002														
2003														
2004							Rc	Rc						
2005				Rc		Rc	Rc	Rc	Rc					
2006			Rc	Rc		Rc	Rc	Rc	Rc					
2007							Rc	Rc	Rc					
2008				Rc			Rc	Rc	Rc					

*Not*e. An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-1 to compare the actual means for each year.

Table A-5. Differences by Year for the MAB FSIQ Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999							Rc	Rc	Rc					
2000														
2001														
2002														
2003														
2004														
2005								Rc						
2006														
2007														
2008								Rc						

Note: An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-1 to compare the actual means for each year.

Table A-6. Differences by Year for the NEO-PI-R Neuroticism Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999														
2000														
2001														
2002					Rc	Rc						Rc	Rc	Rc
2003						Rc						Rc	Rc	Rc
2004														
2005														
2006														
2007														
2008														

Note: An "Rc" is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row "R" mean larger than the lower case "c" column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-1 to compare the actual means for each year.

Table A-7. Differences by Year for the NEO-PI-R Openness to Experience Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999														
2000														
2001											Rc			
2002											Rc			
2003														
2004														
2005														
2006														
2007														
2008														

Note: An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-1 to compare the actual means for each year.

Table A-8. Differences by Year for the NEO-PI-R Agreeableness Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998										Rc				
1999		Rc							Rc	Rc				
2000									Rc	Rc				
2001										Rc				
2002														
2003														
2004														
2005														
2006														
2007										Rc				
2008		Rc							Rc	Rc				

Note: An "Rc" is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row "R" mean larger than the lower case "c" column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table 15\\A-1 to compare the actual means for each year.

Table A-9. Differences by Year for NEO-PI-R Conscientiousness Score: Primary Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998								Rc						
1999														
2000														
2001														
2002														
2003														
2004														
2005								Rc	Rc					
2006								Rc	Rc					
2007							Rc	Rc	Rc					
2008								Rc	Rc					

Note: An "Rc" is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row "R" mean larger than the lower case "c" column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-1 to compare the actual means for each year.

Table A-10. Correlation of MAB–II IQ Scores with C-Score Percentile Rank by Year for Primary Training

Year	VIQ		PIQ		FSIQ	
	U	C	U	C	U	C
1995[a]	.03	-----	-.30	-----	-.17	-----
1996	.07	-.12	.04	-.02	.06	.12
1997	.01	.08	.19*	.25	.12*	.25
1998	.04	.15	.15*	.25	.12*	.26
1999	.03	.12	.15*	.24	.11*	.24
2000	.11*	.23	.11*	.23	.13*	.29
2001	.12*	.32	.16*	.34	.17*	.37
2002	.06	.18	.16*	.27	.13*	.27
2003	.11*	.26	.12*	.27	.14*	.30
2004	.21*	.46	.18*	.40	.22*	.44
2005	.09	.24	.06	.19	.09	.21
2006	.16*	.41	.14*	.38	.18*	.41
2007	.08*	.19	.04	.15	.07*	.16
2008	.13*	.32	.11*	.29	.14*	.32

Note. U means uncorrected and C means corrected for range restriction (Thorndike, 1949).
[a]The correlations for 1995 were not corrected due to the small sample size (N = 17).
*$p < .05$

Table A-11. Correlation of NEO PI–R Domain Scores with C-Score Rank by Year for Primary Training

	N		E		O		A		C	
Year	U	C	U	C	U	C	U	C	U	C
1995[a]	-.23	-----	.21	-----	-.31	-----	.04	-----	.02	-----
1996	-.01	.00	.11	.12	-.10	-.12	-.12	-.09	.04	.05
1997	-.02	-.02	-.01	-.04	-.08	-.01	-.07	-.08	.01	-.03
1998	-.09*	-.10	.04	-.01	-.12*	-.04	.04	.02	.04	.02
1999	-.05	-.07	-.05	-.09	-.01	.05	-.02	-.02	.04	.03
2000	-.03	-.03	.04	-.01	-.10*	.00	-.01	-.01	.06	.03
2001	-.03	-.02	-.01	-.10	-.03	.13	-.05	-.07	.00	-.05
2002	-.02	-.04	.06	.01	-.13*	-.04	-.05	-.05	-.02	-.04
2003	-.01	-04	.05	-.02	-.05	.08	-.04	-.05	-.08*	.04
2004	.04	.06	-.10*	-.27	.03	.22	-.04	-.06	-.02	-.10
2005	.02	-.02	.03	-.03	-.05	.09	.02	-.02	.07	.08
2006	.04	-.03	-.04	-.13	-.07	-.07	-.05	-.09	.10*	-.08
2007	.02	-.02	.06	.00	-.12*	-.02	.04	.03	.00	.00
2008	-.07*	-.09	-.05	-.18	-.02	.12	-.03	-.02	.03	-.05

Note. U means uncorrected and C means corrected for range restriction (Thorndike, 1949).
[a]The correlations for 1995 were not corrected due to the small sample size (N = 17).
*$p < .05$

Table A-12. Descriptive Statistics by Year for T-38 Training

		C-score		VIQ		PIQ		FSIQ		N		E		O		A		C	
Year	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1995	10*	0.28	0.32	121.30	9.90	118.90	14.41	121.80	11.72	40.50	9.16	61.70	5.72	55.70	9.88	51.30	9.80	59.00	6.93
1996	77	0.47	0.29	118.51	6.35	120.34	8.80	120.86	6.33	44.25	8.26	58.05	8.83	50.08	9.80	38.73	11.36	56.12	9.09
1997	149	0.54	0.30	119.29	6.30	119.83	8.17	120.96	6.26	46.24	9.96	57.91	9.67	50.60	10.87	41.19	10.18	54.40	10.95
1998	289	0.47	0.29	120.09	6.67	120.27	8.35	121.58	6.70	46.18	9.49	57.55	10.15	50.96	10.37	43.79	11.85	55.95	9.95
1999	302	0.45	0.30	121.13	6.71	121.32	7.82	122.74	6.55	45.74	9.31	56.82	9.98	51.16	11.27	43.77	10.73	54.42	9.90
2000	357	0.43	0.28	120.12	6.37	119.69	7.89	121.33	6.32	45.42	8.98	56.78	9.15	50.25	11.11	44.39	9.85	55.77	9.88
2001	379	0.40	0.29	120.60	5.82	119.64	7.54	121.55	5.89	47.07	9.79	57.63	9.37	51.63	9.58	43.51	11.42	54.84	10.35
2002	359	0.51	0.28	120.22	6.28	119.60	8.27	121.33	6.63	47.45	9.71	58.55	9.40	50.67	10.73	42.14	10.19	53.77	10.78
2003	345	0.54	0.29	120.31	6.80	120.17	7.99	121.66	6.63	47.99	9.98	58.84	9.57	51.14	10.55	41.61	10.53	54.32	10.09
2004	157	0.56	0.27	120.82	6.40	121.73	8.13	122.73	6.44	45.36	9.48	58.13	9.91	50.12	10.85	40.13	9.86	56.28	9.11
2005	179	0.55	0.28	119.21	6.31	122.56	7.75	122.32	6.26	44.86	10.34	60.39	8.96	49.27	8.99	39.99	11.29	55.95	9.62
2006	166	0.50	0.28	120.28	5.87	122.81	6.51	122.98	5.44	45.05	7.97	58.99	10.05	49.99	9.43	41.40	9.10	58.22	9.30
2007	233	0.49	0.30	119.65	5.55	121.15	7.46	121.82	5.57	44.78	8.59	59.40	10.06	49.29	10.52	43.59	10.52	56.95	9.42
2008	295	0.52	0.31	120.07	6.01	121.51	7.23	122.20	5.78	45.09	9.15	58.05	9.89	49.18	9.66	44.01	10.51	57.12	9.84

Table A-13. Differences by Year for the C-Score: T-38 Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997						Rc	Rc							
1998														
1999														
2000														
2001													Rc	
2002							Rc							
2003						Rc	Rc							
2004						Rc	Rc							
2005						Rc	Rc							
2006														
2007														
2008							Rc							

Note: An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-12 to compare the actual means for each year.

Table A-14. Differences by Year for the MAB PIQ Score: T-38 Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999														
2000														
2001														
2002														
2003														
2004														
2005						Rc	Rc	Rc						
2006						Rc	Rc	Rc	Rc					
2007														
2008														

Note: An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-12 to compare the actual means for each year.

Table A-15. Differences by Year for the NEO-PI-R Neuroticism Score: T-38 Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999														
2000														
2001														
2002														
2003													Rc	
2004														
2005														
2006														
2007														
2008														

Note: An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-12 to compare the actual means for each year.

Table A-16. Differences by Year for the NEO-PI-R Extraversion Score: T-38 Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999														
2000														
2001														
2002														
2003														
2004														
2005					Rc	Rc								
2006														
2007														
2008														

Note: An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-12 to compare the actual means for each year.

Table A-17. Differences by Year for the NEO-PI-R Agreeableness Score: T-38 Training

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999														
2000		Rc								Rc	Rc			
2001														
2002														
2003														
2004														
2005														
2006														
2007														
2008														

Note: An “Rc” is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row “R” mean larger than the lower case “c” column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-12 to compare the actual means for each year.

Table A-18. Differences by Year for the NEO-PI-R Conscientiousness Score: T-38 Training

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1995														
1996														
1997														
1998														
1999														
2000														
2001														
2002														
2003														
2004														
2005														
2006					Rc			Rc	Rc					
2007														
2008														

Note: An "Rc" is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row "R" mean larger than the lower case "c" column mean. If there is no significant difference the cell is empty. The reader is also referred to the means in Table A-12 to compare the actual means for each year.

Table A-19. Correlation of MAB-II IQ Scores with C-Score Percentile Rank by Year for T-38 Training

	VIQ		PIQ		FSIQ	
Year	U	C	U	C	U	C
1995[a]	.23	-----	.44	-----	.39	-----
1996	.19	.42	.06	.30	.15	.33
1997	.04	.18	.16*	.28	.13	.30
1998	.12*	.29	.14*	.30	.15*	.32
1999	-.03	.00	.14*	.19	.06	.13
2000	.20*	.46	.17*	.41	.22*	.47
2001	.08	.24	.13*	.28	.13*	.31
2002	.08	-.01	.00	-.06	.04	.09
2003	.08	.22	.15*	.29	.14*	.30
2004	.19*	.43	.13	.35	.19*	.41
2005	-.01	-.02	-.01	-.02	-.02	.00
2006	.08	.21	.04	.16	.07	.19
2007	.16*	.41	.05	.29	.13	.33
2008	.29*	.60	.09	.41	.23*	.52

Note: U means uncorrected and C means corrected for range restriction (Thorndike, 1949).
[a]The correlations for 1995 were not corrected due to the small sample size (N = 10).
**p*< .05

Table A-20. Correlation of NEO PI–R Domain Scores with C-Score Percentile Rank by Year for T-38 Training

	N		E		O		A		C	
Year	U	C	U	C	U	C	U	C	U	C
1995[a]	.49	-----	-.25	-----	-.17	-----	-.21	-----	-.24	-----
1996	-.07	-.21	.08	.04	-.06	.02	.05	-.05	.03	.07
1997	.09	.03	.10	.05	-.23*	-.15	-.25*	-.26	.06	.03
1998	-.08	-.13	-.01	-.07	-.06	-.09	-.01	.00	.14*	.15
1999	-.07	-.08	.08	.05	.05	.06	.02	.04	.08	.09
2000	.02	.02	-.04	-.21	-.05	.12	-.01	-.10	.10	.07
2001	-.02	-.04	.01	-.09	-.14*	.00	-.05	-.08	.10*	-.07
2002	.04	.05	.14*	.15	-.03	-.06	-.00	-.01	-.06	-.05
2003	.02	-.03	.04	-.02	-.05	.08	-.02	-.02	.08	.05
2004	.05	.13	.10	-.12	-.11	.13	-.01	-.07	-.05	-.03
2005	-.01	.00	.03	.04	-.01	-.07	-.10	-.09	.01	.00
2006	.01	-.04	-.10	-.12	-.06	.01	.00	.03	-.02	.06
2007	.10	.02	.05	-.06	-.03	.16	-.09	.01	-.02	.07
2008	.08	.03	-.08	-.33	-.05	.21	-.10	.05	-.01	-.11

Note: U means uncorrected and C means corrected for range restriction (Thorndike, 1949).
[a]The correlations for 1995 were not corrected due to the small sample size (N = 10).
*$p < .05$

Table A-21. Descriptive Statistics by Year for Advanced T-1 Training

		C-score		VIQ		PIQ		FSIQ		N		E		O		A		C	
Year	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2005	297	0.49	0.30	117.94	5.81	120.39	7.72	120.55	5.94	45.44	9.58	58.01	9.70	49.70	10.06	42.78	10.85	55.94	10.17
2006	340	0.50	0.30	117.66	6.23	120.16	7.90	120.22	6.03	46.48	8.32	58.22	9.07	49.76	9.46	43.91	9.46	55.01	9.27
2007	448	0.58	0.28	118.30	6.52	120.24	7.47	120.60	6.11	46.08	8.78	57.45	9.45	49.84	9.78	44.48	10.23	56.29	9.64
2008	439	0.57	0.29	118.67	6.62	120.29	7.97	120.80	6.49	46.92	9.86	57.17	9.64	50.73	9.40	45.72	10.40	55.14	10.27

Note: T-1 Training did not exist prior to 2005.

Table A-22. Differences by Year for the C-Score: T-1 Training

Year	2005	2006	2007	2008
2005				
2006				
2007	Rc	Rc		
2008	Rc			

Note: An "Rc" is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row "R" mean larger than the lower case "c" column mean. If there was no significant difference the cell is empty. The reader is also referred to the means in Table A-21 to compare the actual means for each year.

Table A-23. Differences by Year for the NEO-PI-R Agreeableness Score: T-1 Training

Year	2005	2006	2007	2008
2005				Rc
2006				
2007				
2008				

Note: An "Rc" is placed in each cell of the matrix indicating a significant difference between the row year and column year means for that cell, with the upper case row "R" mean larger than the lower case "c" column mean. If there was no significant difference the cell is empty. The reader is also referred to the means in Table A-21 to compare the actual means for each year.

Table A-24. Correlation of MAB–II IQ Scores with C-Score Percentile Rank by Year for T-1 Training

Score	VIQ		PIQ		FSIQ	
	U	C	U	C	U	C
2005	.14*	.34	.05	.24	.12*	.29
2006	.15*	.36	.07	.28	.13*	.31
2007	.08	.23	.10*	.24	.10*	.24
2008	.06	.13	.01	.07	.03	.06

Note. U means uncorrected and C means corrected for range restriction (Thormdike, 1949).
*p < .05

Table A-25. Correlation of NEO PI–R Domain Scores with C-Score Percentile Rank by Year for T-1 Training

Year	N		E		O		A		C	
	U	C	U	C	U	C	U	C	U	C
2005	.00	.02	.07	-.03	-.12*	.04	.06	.09	.10	.08
2006	.13*	.06	.03	-.05	-.02	.07	-.09	-.08	.13*	.04
2007	.04	-.02	-.03	-.10	-.08	.04	.09	.04	.10*	.09
2008	-.08	-.09	-.04	-.09	.02	.06	.01	.02	.10*	.07

Note. U means uncorrected and C means corrected for range restriction (Thorndike, 1949).
*p < .05

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS

AETC	Air Education and Training Command
AFOQT	Air Force Officer Qualifying Test
ANOVA	Analysis of Variance
d	Standardized difference between 3 means
ENJJPT	Euro-NATO Joint-Jet Pilot Training
GPA	Grade Point Average
M	Mean
MAB	Multidimensional Aptitude Battery
MAB FSIQ	Multidimensional Aptitude Battery Full-Scale IQ
MAB PIQ	Multidimensional Aptitude Battery Performance IQ
MAB VIQ	Multidimensional Aptitude Battery Verbal IQ
MFS	Medical Flight Screening
NEO-PI-R	NEO Personality Inventory-Revised
OTS	Officer Training School
PCSM	Pilot Candidate Selection Method
r	Pearson correlation coefficient
ROTC	Reserve Officer Training Corps
SD	Standard deviation
SUPT	Specialized Undergraduate Pilot Training
TBAS	Test of Basic Aviation Skills
WAIS-R	Wechsler Adult Intelligence Scale – Revised