AFHRL-TP-88-23

(2)

AIR FORCE HUMAN RESOURCES LABORATORY

PERSONALITY, ATTITUDES, AND PILOT TRAINING PERFORMANCE: FINAL ANALYSIS

Thomas R. Carretta
Frederick M. Siem

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

October 1988

Interim Technical Paper for Period September 1983 - December 1987

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

DANIEL L. LEIGHTON, Lt Col, USAF
Chief, Manpower and Personnel Division

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-88-23	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Manpower and Personnel Division 6b. OFFICE SYMBOL (If applicable) AFHRL/MOEA	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (If applicable) HQ AFHRL	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62703F; PROJECT NO. 7719; TASK NO. 18; WORK UNIT ACCESSION NO. 45
11. TITLE (Include Security Classification) Personality, Attitudes, and Pilot Training Performance: Final Analysis	
12. PERSONAL AUTHOR(S) Carretta, T.R. & Siem, F.M.	
13a. TYPE OF REPORT Interim 13b. TIME COVERED FROM Sep 83 TO Dec 87	14. DATE OF REPORT (Year, Month, Day) October 1988 15. PAGE COUNT 30
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
05	08	
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Air Force Officer Qualifying Test (AFOQT); personality; aircrew selection; pilot selection; Basic Attributes Tests (BAT)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Developments in research concerning personality characteristics have led to a renewed interest in applications of individual differences measures for selection of pilot candidates. Recent research efforts have focused on selecting for positive characteristics, rather than screening out pathological traits. Another development is the use of tests in which the dimension of interest is not readily apparent to the test taker.

In the present investigation, five personality and attitudinal tests were administered to United States Air Force (USAF) pilot candidates as part of an experimental test battery under consideration for operational use in pilot selection and classification, the Basic Attributes Tests (BAT) System. These tests were designed to assess decisiveness, risk-taking, self-confidence, survival attitudes, and field dependence-independence. Scores from these tests were examined for their utility in predicting final training outcome (graduation or elimination) and a follow-on training assignment (fighter or non-fighter aircraft).

(Continued)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b. TELEPHONE (Include Area Code) (512) 536-3877 22c OFFICE SYMBOL AFHRL/TSR

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Item 19 (Concluded):

Results indicated that as a group the tests demonstrated weak relationships with the performance criteria. No test was valid against both performance outcomes. Measures from all five tests were combined into a model that also included scores from the Air Force Officer Qualifying Test (AFOQT), the paper-and-pencil examination currently used for USAF pilot selection. Only the test of self-confidence appeared to contribute unique variance in predicting successful completion of pilot training, over and above that explained by the AFOQT.

AFHRL Technical Paper 88-23 October 1988

PERSONALITY, ATTITUDES, AND PILOT TRAINING PERFORMANCE: FINAL ANALYSIS

Thomas R. Carretta
Frederick M. Siem

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed and submitted for publication by

William R. Ercoline, Lt Col, USAF
Chief, Cognitive Skills Assessment Branch

SUMMARY

United States Air Force (USAF) pilot candidates were administered a computerized test battery, the Basic Attributes Tests (BAT), that is currently being validated for use in pilot selection and classification. Included in the battery were five tests measuring personality and attitudinal characteristics. These tests were evaluated singly and in combination in terms of their ability to enhance the prediction of pilot training outcomes, relative to that prediction offered by the paper-and-pencil measures being used operationally. Based on results from the present data, it was recommended that four of the five tests under review be eliminated from the BAT and that other measures of personality and attitudinal characteristics be evaluated for possible inclusion in a subsequent version of the BAT battery.

PREFACE

This work was completed under Work Unit 77191845 in support of a Request for Personnel Research (RPR 78-11, Selection for Pilot Training) submitted by Air Force training program managers. This paper is intended to serve as interim documentation regarding the personality/attitudinal tests of the Basic Attributes Tests (BAT) battery.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

PERSONALITY, ATTITUDES, AND PILOT TRAINING PERFORMANCE: FINAL ANALYSIS

I. INTRODUCTION

Most research into military pilot selection and classification has concentrated on psychomotor skills and perceptual/cognitive abilities (e.g., Imhoff & Levine, 1981). Relationships among pilot personality, attitudes, and performance have been researched less, although interest in the topic dates back to World War I (North & Griffin, 1977). The present technical paper focuses on recent efforts to validate a number of personality and attitude measures included in a computerized battery of tests currently being evaluated by the United States Air Force (USAF) called the Basic Attributes Tests (BAT) battery.

Although in the past 50 years several studies have explored relationships between pilot characteristics and performance, there has been little progress in using measures of individual differences to predict aviator training and performance criteria (Griffin & Mosko, 1977). Two associated factors may account for the weak relationship between personality tests and outcome criteria. One is that the tests generally have focused on distinguishing between normal and abnormal individuals. The second factor is that such tests have been prone to response bias; that is, subjects guess what the test is designed to measure and fake their responses accordingly.

Recent developments in personality testing have addressed both of these issues. One development has been the design of tests in which the dimension being measured is not immediately apparent. A number of these measures have been used in the Air Force (Mullins, 1960, 1962), such as Dot Estimation and Self-Crediting Word Knowledge.

Another development in personality testing has been the design of tests in which the response alternatives to items are equivalent in terms of social desirability, minimizing the tendency of subjects to fake their responses (North & Griffin, 1977). The Activities Interest Inventory, for example, requires the subject to choose between two activities which differ only in the degree of riskiness associated with those activities.

A third development is the increasing use of personality tests to select for positive attributes, as opposed to screening for possible pathological attributes. Helmreich and his colleagues, for example, have found that among both airline and general aviation pilots the characteristics of self-assertiveness, interpersonal orientation and achievement motivation are each associated with attitudes and performance (Helmreich, 1982; Siem, 1987; Siem & Helmreich, 1985.)

The five tests described below were selected for inclusion in the BAT battery to measure domains identified as having potential for pilot selection and classification (Imhoff & Levine, 1981). In particular, the tests focus on the measurement of decision-making style, risk-taking attitudes, self-confidence and field dependence/independence (see Table 1). These measures were chosen based on the observation that a pilot, particularly when flying a jet fighter, must analyze accurately situations that involve a high degree of risk and then respond decisively yet without acting impulsively (Imhoff & Levine, 1981).

As their use was intended to improve present USAF pilot selection practices, these personality and attitude measures were assessed here in terms of their ability to explain unique variance in the various criteria in pilot training performance; that is, criterion variance over and above that explained by the currently used selection instruments (subtest scores of the Air Force Officer Qualifying Test [AFOQT]). Because the AFOQT subtests are cognitive/perceptual in nature, it was expected that they would not be correlated highly with the personality/attitudinal measures from the BAT.

Table 1. Basic Attributes Tests (BAT):
Summary of Personality/Attitudinal Tests

Test name	Duration (min.)	Reference	Attributes measured	Measures of interest
Dot Estimation	5	Mullins, 1962	Compulsiveness/decisiveness	Number of trials completed, number of correct responses, time spent on test
Risk-Taking	10	Slovic, 1966	Risk-taking, decision making	Response time, amount of risk taken
Self-Crediting Word Knowledge	10	Mullins, 1962	Self-assessment ability, self-confidence	Response time, response accuracy, subject's prediction of own performance
Activities Interest Inventory	10	Mullins, 1962	Survival attitudes, risk-taking	Response time, number of high-risk choices
Embedded Figures	15	Witkin, 1949	Field dependence/independence	Response time, response accuracy

II. METHOD

Subjects

The subjects in this study were 1,992 USAF officer candidates tested on the Basic Attributes Tests (BAT) battery. As not all BAT-tested subjects were accepted into Undergraduate Pilot Training (UPT) or completed the training, the sample sizes for the various prediction and criterion measures vary (see Table 2). For a definition of criterion measures, see below.

Table 2. Numbers of Subjects Available

Prediction/criterion measures	N
AFOQT BAT Personality Tests	1,992
UPT Outcome (pass/fail)	812
ATRB Rating (TTB/FAR)	534

Instrumentation

AFOQT

The AFOQT is a paper-and-pencil test battery consisting of 16 subtests. Scores from the subtests are combined into five composite measures: Verbal, Quantitative, Academic Aptitude (Verbal and Quantitative combined), Pilot, and Navigator-Technical. See Table 3 for the subtests that make up each AFOQT composite.

Table 3. Composition of AFOQT Form O Aptitude Composites

Subtest	Verbal	Quantitative	Academic aptitude	Pilot	Navigator-technical
Verbal Analogies	X		X	X	
Arithmetic Reasoning		X	X		X
Reading Comprehension	X		X		
Data Interpretation		X	X		X
Word Knowledge	X		X		
Math Knowledge		X	X		X
Mechanical Comprehension				X	X
Electrical Maze				X	X
Scale Reading				X	X
Instrument Comprehension				X	
Block Counting				X	X
Table Reading				X	X
Aviation Information				X	
Rotated Blocks					X
General Science					X
Hidden Figures					X

In the analyses described below, raw scores for the 16 subtests are used rather than composite scores. This was done for two reasons: first, to identify the content areas of the AFOQT that are related most closely to flight training performance; second, to determine whether the BAT personality/attitude tests are able to explain unique variance in flight training performance not accounted for by the 16 AFOQT subtests.

Dot Estimation

The psychological factor assessed by this test is compulsiveness/decisiveness. Two boxes containing an arbitrary number of dots are presented on the screen. One of the two boxes has one more dot than the other. The subject's task is to determine, as quickly as possible, which of the two boxes contains the greater number of dots. The subject is not told to count the dots in each box, but told only to decide as quickly and accurately as possible which has the greater number.

In the present effort, reaction time and accuracy of response were recorded on each trial. This was the only test in the battery that had a fixed time limit (5 minutes, maximum of 55 trials).

Risk-Taking

This test assesses risk-taking tendency in making decisions. Ten boxes are presented in two rows of five boxes each. The subject is told that nine of the ten boxes contain a reward, whereas one of the boxes is a "disaster" box. The subject is allowed to select the boxes one at a time. If the selected boxes contain a payoff, the subject is allowed to keep it; but if the subject chooses the disaster box, all of the payoff earned on that trial is lost. The average number of boxes selected provides an index of the subject's tendency for taking risks when making decisions.

Response time per choice and number of boxes chosen were recorded on each of the 30 trials. Unknown to the subject, during 12 of the 30 trials there was no disaster box (i.e., no risk). This was done to get a clean measure of risk-taking behavior, as performance on the disaster box trials might have been affected by chance.

Self-Crediting Word Knowledge

Self-assessment ability and self-confidence are the psychological attributes measured by this test. This is essentially a vocabulary test where the subject is presented with a "target" word and five other words from which its closest synonym has to be chosen. There are three blocks of ten questions each. The target words become increasingly difficult with each successive block. The subject is informed of this increasing difficulty and is required to make a bet prior to each block which reflects how well he/she expects to perform. Response time and accuracy of response were recorded on each of the 30 trials.

Activities Interest Inventory

The psychological factors underlying this test are survival attitudes and risk-taking tendency. This test is designed to determine the subject's interest in various activities. The subject is presented with 81 pairs of activities and is asked to indicate a preference for each pair. The subject is told to assume that he/she has the necessary ability to perform each activity. The activity pairs force the subject to choose between tasks that differ on threat to physical survival--sometimes subtly, sometimes not. Here, the measures of interest were the number of high-risk options chosen and the average amount of time required to choose between pairs of activities.

Embedded Figures

This test is designed to assess the psychological factor of field dependence/independence. It should be noted that level of field dependence has been treated as a personality characteristic by some researchers and as a perceptual ability by others.

As this test has been examined separately in another paper (Carretta, 1987), it will not be examined in detail here. However, analyses were performed to determine its relationship to the other BAT tests discussed in this paper.

In this test, the subject is presented with a simple geometric figure and two complex figures. The task is to decide which of the two complex figures has the simple figure within it and to indicate a choice by pressing the keypad button corresponding to the figure. Speed and accuracy of response were recorded on each of the 30 trials.

UPT Performance Criteria

UPT final training outcome was scored as a dichotomous variable with Pass = 1 and Fail = 0. Subjects who passed UPT received a recommendation from an Advanced Training Recommendation Board (ATRB) for advanced training leading to an assignment either as a Tanker-Transport-Bomber (TTB) pilot or a Fighter-Attack-Reconnaissance (FAR) pilot (FAR = 1 and TTB = 0).

Apparatus

The BAT apparatus consists of a super-microcomputer built into a self-contained unit with a glare shield and side panels designed to ensure consistency of testing sessions. The subject responds to the various tests using in combination or individually a two-axis joystick on the right side of the apparatus, a single-axis joystick on the left side, and a keypad in the center of the test unit. The keypad includes the numbers 0 to 9, an "ENABLE" key in the center, and a bottom row with "YES" and "NO" keys and two others labeled "S/L" (for same/left responses) and "D/R" (for different/right responses). Figure 1 is a picture of the test apparatus. During a test session, the test administrator's keyboard is stored under the desk of the test apparatus.

The test battery as used in this study consisted of 15 tests lasting about 3 1/2 hours. After a test administrator initiated the system, the test session was self-paced by the subject. The test session included programmed breaks between tests to avoid problems with mental and physical fatigue.

Procedure

Prior to entry into UPT, each subject was administered both the AFOQT and the BAT. Pilot candidates were commissioned through either the Air Force Reserve Officer Training Corps (AFROTC) or the Air Force Officer Training School (OTS). Candidates commissioned through AFROTC took the AFOQT prior to entering college or while an undergraduate. For AFROTC candidates, the BAT was administered during the summer of their junior year in college. For the OTS candidates, the AFOQT was administered after their attainment of a college degree and the BAT was administered at the beginning of their participation in a 2-week Flight Screening Program (FSP).

All candidates took part in the UPT program, which lasts 49 weeks. The ATRB decision was made at the 42nd week of UPT, with final outcome (pass/fail) assigned at the end of the program.

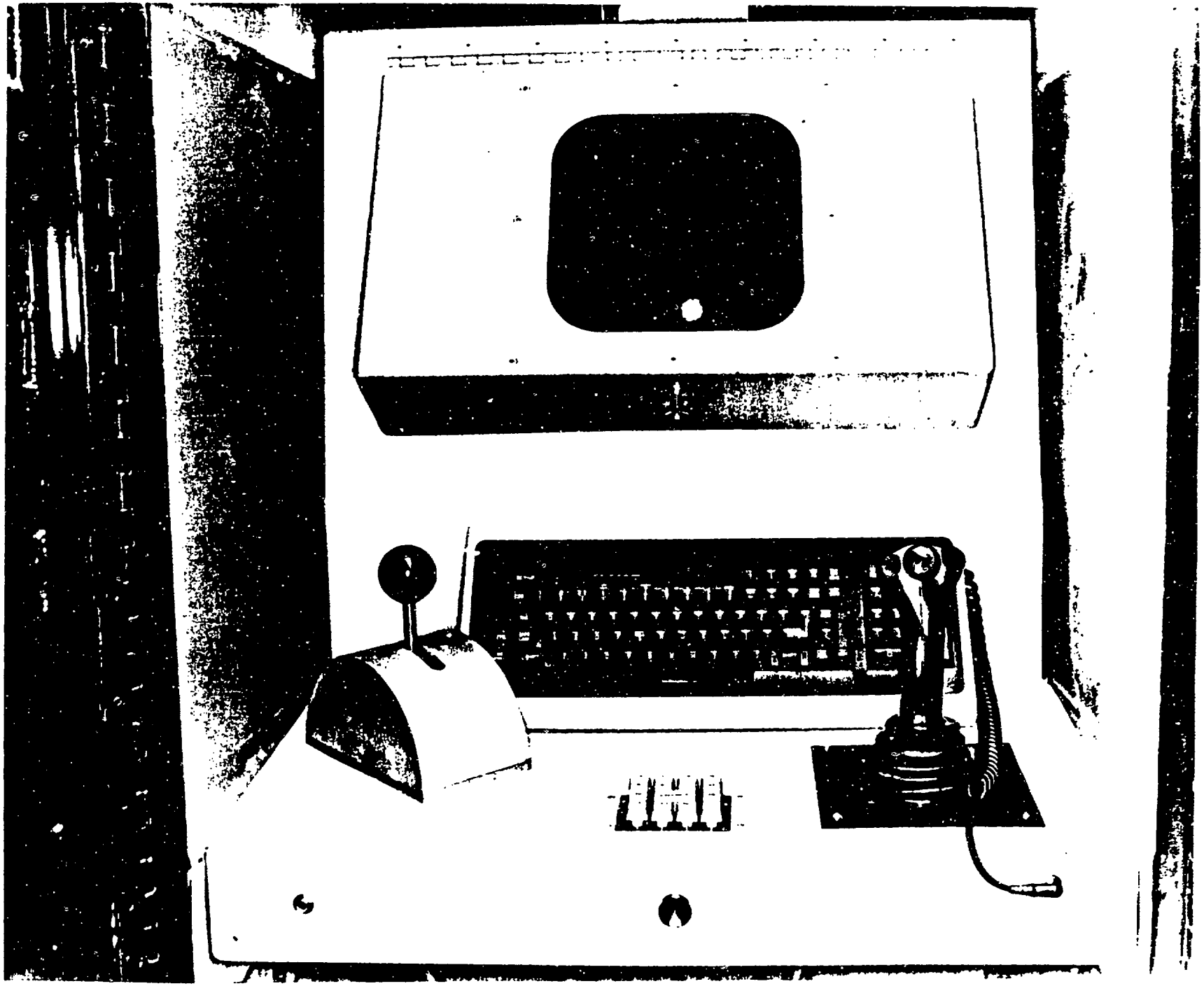

Figure 1. PORTA-BAT Test Unit.

III. RESULTS

AFOQT Scores

A model that used the raw scores from the 16 AFOQT subtests was related significantly to both UPT performance measures. For predicting UPT final outcome (graduation/elimination, $R = .285$, $p \leq .0001$), the subtests that contributed most strongly were Instrument Comprehension ($r = .218$, $p \leq .0001$) and Aviation Information ($r = .173$, $p \leq .01$). Scores on the Rotated Blocks ($r = .102$, $p \leq .10$) and Arithmetic Reasoning subtests ($r = .053$, $p \leq .10$) contributed marginally to prediction of UPT final outcome. Although other subtests had larger zero-order correlations with UPT final outcome than did the Arithmetic Reasoning subtest, they were given less weight in the simultaneous regression model. This suggests that although their zero-order correlations were larger, they were not contributing to the prediction of unique variance in the criterion variable (UPT final outcome). Similar results were obtained for the advanced training recommendation.

For the ATRB recommendation (fighter/non-fighter assignment, $R = .273$, $p \leq .001$), the subtests that contributed significantly were Instrument Comprehension ($r = .155$, $p \leq .05$), Block Counting ($r = -.008$, $p \leq .05$), and Table Reading ($r = .117$, $p \leq .05$). Arithmetic Reasoning ($r = .129$, $p \leq .10$) and Word Knowledge ($r = -.033$, $p \leq .10$) scores contributed marginally to prediction of advanced training recommendation.

These results suggest that the relative importance of the ability domains measured by the 16 AFOQT subtests may change during the course of training. Procedural knowledge about flying (e.g., Aviation Information) acquired before entering UPT may be most important during the early stages of training. Individual differences in procedural knowledge probably decrease during training as level of flying experience increases. During the later stages of training (when the advanced training recommendation is made), individual differences in information processing ability become more important (e.g., Arithmetic Reasoning, Instrument Comprehension, Table Reading). These regression analyses are summarized in Table 4.

Dot Estimation

Descriptive Measures

This test provided several measures to evaluate compulsiveness/decisiveness, including the number of trials completed, number of correct responses, total amount of time spent performing the test, average response time for correct responses, and percent correct.

As can be seen in Table 5, the average number of trials completed was 49.6 out of a maximum of 55. As previously discussed, this test was designed as a "speeded" test; thus, few subjects should have completed all items. On speeded tests, performance is determined, in part, by the number of trials completed. A performance "ceiling" may have occurred with this test as too many subjects completed all items (65%). This could be avoided in the future by either increasing the number of trials or reducing the time limit to a point where few subjects complete all items.

Average number correct (31.7) and percent correct (65.6%) were acceptable, as subjects were not explicitly instructed to count the number of dots in each box before making a choice.

Table 4. AFOQT Subtest Scores: Summary of UPT Outcome Regression Analyses

AFOQT measure	Correlation with AFOQT measures: UPT outcome M = 0.66 (N = 812)	ATRB outcome M = 0.57 (N = 514)
Subtest		
Verbal Analogies	-.044	.040
Arithmetic Reasoning	.053	.129
Reading Comprehension	-.059	.066
Data Interpretation	.031	.114
Word Knowledge	-.088	-.033
Math Knowledge	-.026	.039
Mechanical Comprehension	.024	.098
Electrical Maze	.011	.041
Scale Reading	.031	.095
Instrument Comprehension	.218****	.155*
Block Counting	.075	-.008*
Table Reading	.057	.117*
Aviation Information	.173**	.121
Rotated Blocks	.102	.048
General Science	.002	.022
Hidden Figures	.027	.046
All 16 Subtests (multiple R)	.285****	.273***

Note. Significance levels (*) refer to the unique contribution of a variable in the context of a reduced set of variables which themselves contribute uniquely to the prediction of a criterion. Critical values for zero-order correlations at the .05 level of significance are .069 for N = 800 (UPT) and .088 for N = 500 (ATRB).

*$p \leq .05$.
**$p \leq .01$.
***$p \leq .001$.
****$p \leq .0001$.

Table 5. Dot Estimation: Means and Standard Deviations

Variable	Mean	SD
Number of Trials Completed	49.6	11.8
Number of Correct Responses	31.7	6.9
Percent Correct (%)	65.6	10.3
Total Time (ms.)	1,143,796.1	74,010.6
Average Response Time (ms.) (correct responses)	5,387.6	4,750.1

N = 1,992.

Factor Structure

The inter-item correlation matrix, presented in Table 6, indicates that there was a speed/accuracy tradeoff. As subjects completed more trials, the proportion of correct responses declined (r = -.65). On the other hand, subjects who spent more time on the test had a higher proportion of correct responses on the trials they completed (r = .56).

Table 6. Dot Estimation: Inter-Item Correlation Matrix

	Variable				
	1	2	3	4	5
1. Number of Trials Completed	1.00				
2. Number of Correct Responses	.87	1.00			
3. Percent Correct	-.65	-.23	1.00		
4. Total Time	-.74	-.58	.56	1.00	
5. Average Response Time (correct responses)	-.92	-.83	.56	.87	1.00

N = 1,992.

The factor solution indicated one principal factor that accounted for 75.7% of the total item variance. This suggested that the Dot Estimation test was unidimensional in nature. Results of the factor analysis are presented in Table 7.

Table 7. Dot Estimation: Summary of Factor Analysis

Variable	Communality	Factor loadings I
Number of Trials Completed	.97	-.98
Number of Correct Responses	.60	-.78
Percent Correct	.33	.57
Total Time	.67	.82
Average Response Time (correct responses)	.99	.99

Factor	Eigenvalue	% of total variance	% of explained variance	Cumulative % explained
I	3.57	75.7	100.0	100.0

N = 1,992.

Inferential Measures

A model that used the five Dot Estimation scores was not related significantly to either of the UPT performance measures: UPT final outcome (R = .039, n.s.), ATRB rating (R = .121, n.s.). A combined model that used the 16 AFOQT subtest scores along with the Dot Estimation scores was related statistically to UPT final outcome (R = .287, $p \leq .0001$) and to advanced training assignment (R = .292, $p \leq .001$). In both cases, the combined model failed to improve prediction above that provided by the AFOQT scores alone at the .05 level of probability. A summary of the Dot Estimation regression analyses is provided in Table 8.

Table 8. Dot Estimation: Summary of UPT Outcome Regression Analyses

Predictor measure	Correlation with predictor: UPT outcome (M = 0.66, N = 812)	Correlation with predictor: ATRB outcome (M = 0.57, N = 514)
Dot Estimation Variables		
Number of Trials Completed	-.015	-.037
Number of Correct Responses	-.005	-.002
Percent Correct	.025	.052
Total Time	.012	.065
Average Response Time (correct responses)	.020	.032
Multiple Correlation		
Dot Estimation	.039	.121
16 AFOQT Subtests	.285****	.273***
Combined Mode'	.287****	.292***
R Square Change	.001	.011

***p ≤ .001.
****p ≤ .0001.

Risk-Taking

Descriptive Measures

The most conceptually interesting performance measures on this test were the average number of boxes chosen (i.e., level of risk) and average response time on each trial. Table 9 summarizes level of performance on the "risk" and "no-risk" trials.

Table 9. Risk-Taking: Means and Standard Deviations

Variable	Number of trials	Mean	SD
Number of Boxes Chosen			
Risk	18	4.5	0.8
No Risk	12	6.9	1.3
Average Response Time (ms.)			
Risk	18	2,663.3	1,675.6
No Risk	12	2,232.8	1,608.8

N = 1,992.

Performance on the no-risk trials suggested that these subjects, in general, applied a somewhat risky strategy (average number of boxes chosen = 6.9). An "optimizing" strategy would be to make five choices per trial to maximize rewards in the long term. Reliability estimates were calculated separately for the 18 risk and 12 no-risk trials, as performance on the risk trials was determined, in part, by chance. The number of boxes chosen was much less reliable for

the risk trials (Cronbach's alpha = .520) than for the no-risk trials (Cronbach's alpha = .954). However, average response time per trial was reliable for both risk (Cronbach's alpha = .910) and no-risk trials (Cronbach's alpha = .972).

Factor Structure

The inter-item correlations, presented in Table 10, indicated that the two "riskiness" measures (number of boxes chosen during risk and no-risk trials) were moderately correlated with each other (r = .61) but not with average response time per trial ($-.06 \leq r \leq .01$). The two average response time measures were related strongly to each other (r = .97).

Table 10. Risk-Taking: Inter-Item Correlation Matrix

Variable	Variable 1	2	3	4
Number of Boxes Chosen (risk)	1.00			
Number of Boxes Chosen (no risk)	.61	1.00		
Average Response Time (risk)	-.06	.00	1.00	
Average Response Time (no risk)	-.05	.01	.97	1.00

N = 1,992.

As expected, the factor analysis yielded two factors; namely, response latency and level of risk. The principal factor consisted of the two average response time variables and accounted for 49.4% of the total item variance (61.6% of the "explained" variance). Both of the number of boxes chosen variables loaded on the second factor, which accounted for 39.9% of the total item variance (38.4% of the explained variance). The factor analysis is summarized in Table 11.

Table 11. Risk-Taking: Summary of Factor Analysis

Variable	Communality	Factor loadings I	II
Number of Boxes Chosen (risk)	.61	-.04	.78
Number of Boxes Chosen (no risk)	.61	.02	.78
Average Response Time (risk)	.97	.98	-.03
Average Response Time (no risk)	.97	.98	-.01

Factor	Eigenvalue	% of total variance	% of explained variance	Cumulative % explained
I	1.94	49.4	61.6	61.6
II	1.21	39.9	38.4	100.0

N = 1,992.

Inferential Measures

As with the Dot Estimation model, performance measures from the Risk-Taking test demonstrated poor predictive utility against UPT final outcome (R = .066, n.s.) and advanced training assignment (R = .062, n.s.).

A combined model that used the Risk-Taking measures along with the 16 AFOQT subtest scores was related significantly to UPT final outcome (R = .289, $p \leq .0001$) and advanced training recommendation (R = .282, $p \leq .01$). As with Dot Estimation, the combined model did not improve prediction above that provided by the AFOQT alone. The Risk-Taking regression analyses are summarized in Table 12.

Table 12. Risk-Taking: Summary of UPT Outcome Regression Analyses

	Correlation with predictor	
	UPT outcome	ATRB outcome
Predictor measure	$\bar{M}$ = 0.66 (N = 812)	$\bar{M}$ = 0.57 (N = 514)
Risk-Taking Variables		
Number of Boxes chosen (risk)	-.053	-.024
Number of Boxes chosen (no risk)	-.029	-.013
Average Response Time (risk)	-.029	-.009
Average Response Time (no risk)	-.023	-.023
Multiple correlation		
Risk-Taking	.066	.062
16 AFOQT Subtests	.285****	.273***
Combined Model	.289****	.282**
R Square Change	.002	.005

**$p \leq .01$.
***$p \leq .001$.
****$p \leq .0001$.

Self-Crediting Word Knowledge

Descriptive Measures

As previously mentioned, this test is essentially a vocabulary test designed to measure self-assessment ability and self-confidence. Self-assessment was operationalized as the difference between the subject's expectations (bet) and his/her actual performance (number correct).

As shown in Table 13, subjects' actual performance (67.1% correct) far exceeded their expectations (39.0% correct). Average response time for correct responses was 8.02 seconds. A speed by accuracy interaction term was calculated by multiplying average response time by percent correct and correcting for the means on those variables. As the interaction term is strongly negative, it indicated that subjects who made more correct responses also responded more quickly (i.e., subjects above the mean on one variable tended to be below the mean on the other variable).

Table 13. Self-Crediting Word Knowledge: Means and Standard Deviations

Variable	Mean	SD
Average Response Time (ms.) (correct responses)	8,022.5	1,914.5
Percent Correct	67.1	10.5
Bet	39.0	10.3
Average Response Time x Percent Correct	-3,555.3	24,830.7

Note. The Average Response Time x Percent Correct interaction term was calculated by subtracting the grand mean from each subject's mean for the two variables and then multiplying the two difference scores together ([subject's average response time - grand mean response time] x [subject's percent correct - grand mean percent correct]).

N = 1,992.

Accuracy of response (Cronbach's alpha = .653) and average response time per trial (Cronbach's alpha = .885) demonstrated acceptable reliability.

Factor Structure

A preliminary evaluation of the factor structure of this test resulted in five performance variables. In addition to average response time, percent correct, bet, and the speed/accuracy interaction term, a fifth variable--difference between actual and expected performance (percent correct minus bet)--was calculated. The fifth variable was dropped, however, because it was correlated too strongly with the other variables and resulted in a communality value equal to 1.0.

The inter-item correlations, summarized in Table 14, indicated that the remaining variables were not related strongly to each other. As expected, actual and expected performance were moderately related (r = .33). Average response time was negatively related to actual (r = -.16) and expected (r = -.21) performance. Subjects who were more self-confident (bet more) were more accurate and responded more quickly than did subjects who were less self-confident (bet less).

Table 14. Self-Crediting Word Knowledge: Inter-Item Correlation Matrix

	Variable			
Variable	1	2	3	4
Average Response Time (correct responses)	1.00			
Percent Correct	-.16	1.00		
Bet	-.21	.33	1.00	
Average Response Time x Percent Correct	-.13	-.19	.00	1.00

N = 1,992.

The factor analysis produced two factors which together accounted for 65.6% of the total item variance. The two "accuracy" scores (percent correct and bet) defined the principal factor, while average response time and the speed/accuracy interaction term defined the second factor.

These two factors reflected the crucial components of this test; namely, accuracy/self-confidence and response speed. Results of the factor analysis are summarized in Table 15.

Table 15. Self-Crediting Word Knowledge: Summary of Factor Analysis

Variable	Communality	Factor loadings I	Factor loadings II
Average Response Time (correct responses)	.21	-.39	-.24
Percent Correct	.49	.62	.32
Bet	.28	.53	.01
Average Response Time x Percent Correct	.32	-.01	.56

Factor	Eigenvalue	% of total variance	% of explained variance	Cumulative % explained
I	0.85	36.9	65.6	65.6
II	0.45	28.7	34.4	100.0

N = 1,992.

Inferential Measures

The Self-Crediting Word Knowledge model was related statistically to UPT final outcome (R = .157, $p \leq .001$) but not to advanced training recommendation (R = .036, n.s.). Contrary to expectations, subjects who took longer to respond were more likely to pass UPT (R = .141, $p \leq .001$). Those who took longer to respond may have been showing caution rather than a lack of confidence.

A combined model that used the 16 subtest scores from the AFOQT along with the scores from the Self-Crediting Word Knowledge test was related statistically to UPT final outcome (R = .312, $p \leq .0001$), and significantly improved prediction above that provided by the 16 AFOQT subtests alone ($F[4,791] = 3.53$, $p \leq .01$). For the ATRB outcome, the combined model showed little improvement over the AFOQT scores alone. Table 16 provides a summary of these regression analyses.

Activities Interest Inventory

Descriptive Measures

As with Risk-Taking, this test was designed to assess attitudes toward risk-taking. The primary measure of interest was the number of high-risk activities chosen by each subject from the activity pairs.

The average number of high-risk activities chosen was 49.6 out of 81 (61.2%). Average response time per trial was 4.48 seconds. The number of high-risk activities chosen and average response time were not statistically related to each other (r = -.07). The reliabilities of response choice (Cronbach's alpha = .864) and response time (Cronbach's alpha = .954) were acceptable. Table 17 presents the means and standard deviations for these measures. A factor analysis was not performed because there were only two variables.

Inferential Measures

Scores on this test were not statistically related to UPT final outcome (R = .043, n.s.) or advanced training recommendation (R = .061, n.s.).

Table 16. Self-Crediting Word Knowledge: Summary of UPT Outcome Regression Analyses

Predictor measure	Correlation with predictor: UPT outcome M = 0.66 (N = 812)	ATRB outcome M = 0.57 (N = 514)
Self-Crediting Word Knowledge Variables		
Average Response Time (correct responses)	.141***	-.026
Percent correct	-.074	.026
Bet	-.063	.019
Average Response Time X Percent correct	.029	.000
Multiple Correlation		
Self-Crediting Word Knowledge	.157***	.036
16 AFOQT Subtests	.285****	.273***
Combined Model	.312****	.278**
R Square Change	.016**	.003

Note. Significance levels (*) refer to the unique contribution of a variable in the context of a reduced set of variables which themselves contribute uniquely to the prediction of a criterion. Critical values for zero-order correlations at the .05 level of significance are .069 for N = 800 (UPT) and .088 for N = 500 (ATRB).

**p $\leq$ .01.
***p $\leq$ .001.
****p $\leq$ .0001.

Table 17. Activities Interest Inventory: Means and Standard Deviations

Variable	Mean	SD
Number of High-Risk Activities Chosen	49.6	9.9
Average Response Time per Trial	4,483.8	1,080.3

N = 1,992.

A combined model that used the 16 AFOQT subtest scores along with the Activities Interest Inventory scores was related statistically to final training outcome (R = .291, $p \leq .0001$) and to advanced training recommendation (R = .276, $p \leq .001$) but did not improve prediction significantly over a model that used only the AFOQT subtests. The Activities Interest Inventory regression analyses are summarized in Table 18.

Table 18. Activities Interest Inventory:
Summary of UPT Outcome Regression Analyses

Predictor measure	Correlation with predictor: UPT outcome M = 0.66 (N = 812)	Correlation with predictor: ATRB outcome M = 0.57 (N = 514)
Activities Interest Inventory Variables		
Number of High-Risk Activities Chosen	-.020	.049
Average Response Time	-.036	-.042
Multiple Correlation		
Activities Interest Inventory	.043	.061
16 AFOQT Subtests	.285****	.273***
Combined Model	.291****	.276***
R Square Change	.003	.002

***$p \leq .001$.
****$p \leq .0001$.

Embedded Figures

Descriptive Measures

The most important performance measures on this test were accuracy of response and average response time. Although overall accuracy of response was acceptable (65.5% correct), accuracy fell below "chance level" (50%) on 11 of the 30 trials. Most of these trials exhibited low correlations with the item-total score, suggesting that the stimuli used on these trials were poor discriminators of performance and should be eliminated from this test. Despite this problem, responses were fairly reliable (Cronbach's alpha = .702).

Average response time for correct responses was 12.2 seconds and was very reliable (Cronbach's alpha = .915). These descriptive measures are summarized in Table 19.

Table 19. Embedded Figures:
Means and Standard Deviations

Variable	Mean	SD
Average Response Time (ms) (correct responses)	12,200.0	4,802.9
Percent Correct (%)	65.5	14.5

N = 1,992.

Additional details regarding the items of this test (e.g., item-total correlations, inter-item correlations, and factor structure) are not discussed here but are available in an earlier paper (Carretta, 1987). Carretta (1987) suggested that performance on the Embedded Figures test could be summarized by three variables: average response time, accuracy of response, and a response time by accuracy interaction term.

Inferential Measures

The Embedded Figures model (average response time, percent correct, and response time by percent correct interaction term) demonstrated poor predictive utility against both of the UPT performance criteria. The model was not statistically related to UPT pass/fail outcome (R = .050, n.s.) or to advanced training recommendation (R = .089, n.s.).

When the Embedded Figures model was combined with the 16 AFOQT subtest scores, the combined model was related statistically to both UPT final outcome (R = .296, $p \leq .0001$) and advanced training recommendation (R = .293, $p \leq .001$). The combined model, however, did not significantly improve prediction of performance above that provided by the AFOQT scores alone. Results from these regression analyses are presented in Table 20.

Table 20. Embedded Figures: Summary of UPT Outcome Regression Analyses

Predictor Measure	Correlation with predictor: UPT outcome M = 0.66 (N = 812)	ATRB outcome M = 0.57 (N = 514)
Embedded Figures Measures		
Average Response Time	-.005	-.020
Percent Correct	-.046	.039
Average Response Time X Percent Correct	-.016	.075
Multiple Correlation		
Embedded Figures	.050	.089
16 AFOQT Subtests	.285****	.273***
Combined Model	.296****	.293***
R Square Change	.006	.011

***p $\leq$ .001.
****p $\leq$ .0001.

Integrated Model

Factor Structure

A factor analysis was performed using the 18 variables from the five tests in order to determine the relationships among them. An examination of the inter-item correlation matrix presented in Table 21 reveals that there are few strong correlations between variables from different tests. This suggests that there is little overlap among the tests in the characteristics being measured. Although the variables within Dot Estimation and, to a lesser extent, Risk-Taking demonstrated good internal consistency, those from the other three tests did not. The Self-Crediting Word Knowledge, Activities Interest Inventory, and Embedded Figures tests lacked clear factor relationships.

Table 21. BAT Personality/Attitudinal Tests: Inter-Item Correlation Matrix

Variable	Variable 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Dot Estimation																		
N Trials Completed	1.00																	
N Correct Responses	.87	1.00																
Total Time	-.74	-.58	1.00															
Average Response Time	-.92	-.83	.87	1.00														
Percent Correct	-.65	-.23	.56	.56	1.00													
Risk-Taking																		
N Boxes Chosen (risk)	.03	.03	-.04	-.05	-.01	1.00												
N Boxes Chosen (no risk)	.04	.03	-.05	-.06	-.03	.61	1.00											
Average Response Time (risk)	.01	.02	.06	.03	.00	-.06	.00	1.00										
Average Response Time (no risk)	.01	.02	.05	.02	.00	-.05	.01	.97	1.00									
Self-Crediting Word Knowledge																		
Average Response Time	-.15	-.13	.19	.18	.12	-.05	-.06	.09	.06	1.00								
Percent Correct	.03	.03	.01	-.01	.02	.02	.01	.02	.02	-.16	1.00							
RT by % Correct	-.01	.00	.02	.01	.02	.01	.03	-.01	.00	-.13	-.19	1.00						
Bet	-.02	-.01	-.01	.01	.03	.05	.04	-.23	-.24	-.21	.33	.00	1.00					
Activities Interest Inventory																		
N High-Risk Choices	.00	.00	-.03	-.01	.02	.06	.06	-.05	-.05	-.04	.06	.01	.12	1.00				
Average Response Time	-.10	-.06	.14	.11	.10	-.06	-.08	.13	.12	.41	-.12	-.01	-.12	-.07	1.00			
Embedded Figures																		
Average Response Time	-.14	-.11	.23	.19	.11	.02	.04	.05	.03	.18	.00	-.03	-.07	-.07	.11	1.00		
Percent Correct	-.02	.02	.06	.04	.09	-.03	-.05	.03	.03	-.02	.11	-.01	.06	.07	.10	.07	1.00	
RT by % Correct	.03	-.01	-.03	-.03	-.06	.02	.02	-.04	-.04	.01	-.02	.04	-.01	-.01	-.02	-.01	-.15	1.00

The factor analysis, presented in Table 22, produced a six-factor solution that accounted for 65.1% of the total item variance. Only factor loadings with a magnitude of .30 or higher are presented, in order to simplify the table. The principal factor can be interpreted as "speededness" or "compulsiveness" as the five variables from Dot Estimation were the only ones that loaded on it. Factors II (response latency) and III (riskiness) were defined by variables from Risk-Taking. Contrary to expectations, the Activities Interest Inventory variables did not cluster with those from Risk-Taking, although both tests were designed to assess attitudes toward risk-taking. The remaining three factors were uninterpretable as each was defined by only two or three variables and, as a result, lacked stability.

Table 22. BAT Personality/Attitudinal Tests: Summary of Factor Analysis

Variable	Communality	Factor loadings I	II	III	IV	V	VI
Dot Estimation							
N Trials Completed	.94	-.95					
N Correct Responses	.99	-.93				.36	
Total Time	.70	.76				.30	
Average Response Time	.97	.97					
Percent Correct	.67	.57				.60	
Risk-Taking							
N Boxes Chosen (risk)	.57			.75			
N Boxes Chosen (no risk)	.64			.80			
Average Response Time (risk)	.95		.96				
Average Response Time (no risk)	.98		.98				
Self-Crediting Word Knowledge							
Average Response Time	.63				.79		
Percent Correct	.61						.71
RT by % Correct	.09						
Bet	.25						.32
Activities Interest Inventory							
N High-Risk Choices	.02						
Average Response Time	.26				.49		
Embedded Figures							
Average Response Time	.09						
Percent Correct	.05						
RT by % Correct	.02						

Factor	Eigenvalue	% of total variance	% of explained variance	Cumulative %
I	3.78	21.8	40.0	40.0
II	2.07	12.4	22.0	62.0
III	1.26	9.1	13.3	75.3
IV	1.04	8.4	11.0	86.3
V	.73	7.2	7.7	94.0
VI	.56	6.2	6.0	100.0

Note. Factor Loadings less than .30 omitted.
N = 1,992.

The goal of this factor analysis was to identify the common and unique variance among the 18 variables from the five BAT tests, and to produce a minimum number of meaningful factor scores to be used as predictors of flight training performance. However, because the factor solution was not clear, the original 18 variables, rather than the factor scores, were used in an integrated model.

Inferential Measures

A 34-predictor regression equation that used the 16 AFOQT subtest scores along with the 18 BAT variables was related significantly to UPT final outcome (R = .346, $p \leq .0001$). This model was compared to a reduced model that also was related significantly to UPT final outcome (AFOQT subtests and Self-Crediting Word Knowledge scores, R = .312, $p \leq .0001$). The two models did not differ significantly in their predictive utilities ($F[14,777]$ = 1.41, n.s.). That is, scores from the Dot Estimation, Risk-Taking, Activities Interest Inventory, and Embedded Figures tests did not significantly improve the prediction of successful completion of pilot training above that provided by the AFOQT subtests and Self-Crediting Word Knowledge scores. The 34-predictor AFOQT/5 BAT test model was related significantly to advanced training recommendation (R = .326, $p \leq .01$) but did not significantly improve prediction above that provided by the 16 AFOQT subtests alone (F [18,499] = 0.98, n.s.).

Summary

The AFOQT subtest scores as a group demonstrated a moderately strong relationship with UPT performance. It should be noted that the relative importance of the 16 subtests differed for the two flying training outcome measures.

The five sets of personality measures from the BAT were sufficiently reliable to be used in selection systems; however, none of the BAT tests was related statistically to both UPT final outcome and advanced training recommendation. Performance on the Self-Crediting Word Knowledge test was related to UPT final outcome. Subjects who took longer to respond (i.e., were more cautious) were more likely to complete training successfully.

IV. DISCUSSION

There are several explanations for the poor predictive utility demonstrated by these personality/attitudinal tests. One explanation is that the BAT tests may not be measuring the characteristics they were designed to measure (i.e., poor construct validity). Although each test was adapted from a previously validated paper-and-pencil test, no subjects were given both the BAT and the paper-and-pencil versions of the tests. As a result, the BAT tests can be evaluated in terms of face validity, but not construct validity.

Even if the BAT tests have acceptable construct validity, scores on them were not found to be related strongly to pilot training performance. Subjects in this study may have been too similar to one another in terms of the characteristics measured by these tests, or they may have been faking their responses to present a positive image to others, or their "true" personalities may not have emerged because of situational pressures. Another possible explanation is that a "personality/attitudinal profile" that considered several characteristics together might be related more closely to training performance than would any single characteristic alone. Although the personality/attitudinal profile hypothesis was not supported by results from the integrated model, this does not mean that personality and attitudes are not related to flying training performance or that research with personality/attitudinal measures should be abandoned.

Recent efforts by Spence and others (e.g., Spence & Helmreich, 1983; Spence, Helmreich, & Holahan, 1979) have yielded promising relationships among measures of interpersonal skills, need for achievement, and pilot performance. In a research effort being sponsored by the National Aeronautics and Space Administration and the US Navy, other personality attributes not considered here are being evaluated including measures of locus of responsibility (Reid & Ware, 1973), instrumentality and interpersonal orientation (Spence et al. 1979), mastery and competitiveness (Spence & Helmreich, 1983) and other personality factors (Dahlstrom, Welsh, & Dahlstrom, 1972).

V. CONCLUSIONS

Each of the five BAT tests included in this study exhibited acceptable reliability. However, none of them was related statistically to both measures of flying training performance (graduation/elimination, advanced training recommendation). Performance on the Self-Crediting Word Knowledge test was related statistically to UPT final outcome.

As a result, it is suggested that only the Self-Crediting Word Knowledge test be retained in the BAT battery. Future studies are planned to evaluate the construct validity of this test by administering it with other measures of self-confidence and self-assessment.

REFERENCES

Carretta, T.R. (1987). Field dependence-independence and its relationship to flight training performance (AFHRL-TP-87-36, AD-A188 888). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Dahlstrom, W.G., Welsh, G.S., & Dahlstrom, L.E. (1972). An MMPI handbook (Vol. I & II). Minneapolis, MN: University of Minnesota Press.

Griffin, G., & Mosko, J. (1977). Naval aviation attrition 1950-1976: Implications for the development of future research and evaluation (NAMRL 1237). Pensacola FL: Naval Aerospace Medical Research Laboratory.

Helmreich, R.S. (1982). Pilot selection and training. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Imhoff, D.L., & Levine, J.M. (1981). Perceptual-motor and cognitive performance test battery for pilot selection (AFHRL-TR-80-27, AD-A094 317). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Mullins, C.J. (1960). Problems in designing objective personality measures. Paper presented at the Tri-Service Conference on Research Methods and Techniques in Personnel, Washington DC.

Mullins, C.J. (1962). Objective tests of self-confidence (PR-TM-62-6). Lackland AFB, TX: Selection and Classification Branch, Personnel Research Laboratory.

North, R., & Griffin, G. (1977). Aviation selection 1919-1977 (Special Report SR77-2). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Reid, D.W., & Ware, E.E. (1973). Multidimensionality of internal-external control: Implications for past and future research. Canadian Journal of Behavioral Science, 5, 264-271.

Siem, F.M. (1987). The effects of aircrew member personality on social interaction and performance. Unpublished dissertation, University of Texas, Austin, TX.

Siem, F.M., & Helmreich, R.L. (1985). Cockpit attitudes in the general aviation community. Paper presented at the first annual meeting of the Alamo chapter of the Human Factors Society, San Antonio, TX.

Slovic, P. (1966). Risk-taking in children: Age and sex differences. Child Development, 37, 169-176.

Spence, J.T., & Helmreich, R.L. (1983). Achievement-related motives and behavior. In J.T. Spence (Ed.), Achievement and achievement motives: Psychological and sociological approaches. New York: W.H. Freeman & Co.

Spence, J.T., Helmreich, R.L., & Holahan, C.K. (1979). Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. Journal of Personality and Social Psychology, 37, 1673-1682.

Witkin, H.A. (1949). Perception of body position and the position of the visual field. Psychological Monographs, 63, 1-46.