DTIC FILE COPY

AFHRL-TR-89-68

AD-A222 253

AIR FORCE

HUMAN RESOURCES

CROSS-VALIDATION OF EXPERIMENTAL USAF PILOT TRAINING PERFORMANCE MODELS

Thomas R. Carretta

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

May 1990
Interim Technical Report for Period July 1986 - August 1989

Approved for public release; distribution is unlimited.

LABORATORY

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 report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

HAROLD G. JENSEN, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
May 1990

3. REPORT TYPE AND DATES COVERED
Interim - July 1986 to August 1989

4. TITLE AND SUBTITLE
Cross-Validation of Experimental USAF Pilot Training Performance Models

5. FUNDING NUMBERS
PE - 62703F
PR - 7719
TA - 18
WU - 45

6. AUTHOR(S)
Thomas R. Carretta

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

8. PERFORMING ORGANIZATION REPORT NUMBER
AFHRL-TR-89-68

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

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A series of studies have indicated that individual differences in hand-eye coordination, information processing ability, personality and attitudes are related to USAF pilot training performance. The current investigation was designed to cross-validate these results. Eight hundred eighty-five (885) USAF Undergraduate Pilot Training (UPT) students were divided randomly into two groups. Pilot selection models that used a combination of Air Force Officer Qualifying Test (AFOQT) and Basic Attributes Test (BAT) battery scores were developed independently for each group and then cross-validated with the other group.

In the model development phase, subjects with good hand-eye coordination who made quick decisions were more likely to complete UPT successfully in both groups. Although there was some reduction in the validity coefficients in the cross-validation phase, the selection models were related significantly to UPT final outcome in both groups. These results suggest that the AFOQT/BAT pilot selection models are sufficiently robust to be used as adjuncts to operational USAF pilot trainee selection procedures.

Keywords:

14. SUBJECT TERMS
aptitude testing
Basic Attributes Test
pilot selection, (SW)

15. NUMBER OF PAGES
14

16. PRICE CODE

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

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

SUMMARY

The Basic Attributes Test (BAT) battery is a computerized test battery designed to measure individual differences in psychomotor skills, information processing abilities, personality and attitudes. In previous studies, several of these tests have demonstrated utility for supporting USAF pilot selection decisions and adding to the predictive validity of the Air Force Officer Qualifying Test (AFOQT), the ability measure currently included in the USAF pilot selection process.

The purpose of this study was to cross-validate pilot selection models that used a combination of AFOQT and BAT test scores to determine the generalizability of the original prediction models reported in Carretta (1989). To perform the cross-validation, 885 USAF Undergraduate Pilot Training (UPT) students were assigned randomly to two groups. Pilot selection models that used a combination of AFOQT and BAT test scores were developed independently for each group. The "best-fitting" regression weights from each group were then applied to subjects in the other group to determine the generalizability of the regression solutions.

In the model development phase, subjects with good hand-eye coordination who made quick and accurate decisions were more likely to complete UPT successfully in both groups. Although there was some reduction in the validity coefficients in the cross-validation phase, the selection models were related significantly to UPT final outcome in both groups. These results suggest that the AFOQT/BAT pilot selection models are robust enough to be used as adjuncts to operational USAF pilot trainee selection procedures.

PREFACE

This project was conducted under work unit 77191845 in support of Request for Personnel Research (RPR) 78-11, Selection for Undergraduate Pilot Training, issued by Air Training Command.

Appreciation is extended to Mr Robert Levine from OAO, Inc. for his assistance in preparing the data files and programming the data analyses, and to Ms. Gloria Koenig for administrative support. Finally, the author extends thanks to Maj Dave Perry, Dr. Joseph L. Weeks, and Dr. William E. Alley for their comments and technical support during this project.

TABLE OF CONTENTS

LIST OF TABLES

CROSS-VALIDATION OF EXPERIMENTAL USAF
PILOT TRAINING PERFORMANCE MODELS

I. INTRODUCTION

In 1955, the United States Air Force (USAF) discontinued the use of apparatus-based testing as a component of its aircrew selection and classification system. Previously, several testing devices had provided perceptual and motor skill measures that were useful for classifying aircrew applicants into job specialties (pilot versus navigator) and for predicting preliminary flight training outcome (Passey & McLaurin, 1966). Apparatus-based testing was discontinued primarily for administrative reasons including (a) the decision to decentralize the selection process, and (b) the difficulty of keeping the electromechanical testing devices calibrated and the test administration procedures consistent across multiple test sites (McGrevy & Valentine, 1974).

Since then, the variables considered in pilot candidate selection have included medical fitness, academic performance, aptitude test scores, biographical/background data (e.g., age, college activities, type of college degree) and previous flying experience. Despite the demonstrated validity of current USAF pilot candidate selection procedures for predicting training performance, the increasing expense of pilot training and the demands associated with modern aircraft make it crucial that the best-qualified pilot training applicants be selected.

Recently, several studies have demonstrated the utility of a computerized test device known as the Basic Attributes Test (BAT) system for adding to the predictive validity of currently used USAF pilot selection procedures (Bordelon & Kantor, 1986; Carretta, 1989; Kantor & Carretta, 1988). Based on these results, eight tests from the BAT battery are currently under review as adjuncts to the selection process of Undergraduate Pilot Training (UPT) students.

Concerns have been raised regarding the stability of the pilot selection model (Carretta, 1989) because the final selection model was developed using a stepwise regression approach and because there is some evidence of redundancy among the test measures. Therefore, a different final regression solution could occur with a different sample. The present investigation was conducted to cross-validate these results in order to determine the stability of the original prediction model.

II. METHOD

Subjects

The subjects in this study were 885 U.S. Air Force UPT students from the Air Force Reserve Officer Training Corps (AFROTC) and Officer Training School (OTS) who were tested on both the Air Force Officer Qualifying Test (AFOQT) and BAT batteries. Of these subjects, 478 also had been used in the original validation study (Carretta, 1989). All subjects already had been chosen for UPT on the basis of, in part, their AFOQT scores. The BAT battery currently is not part of the operational USAF pilot candidate selection procedure.

Subjects ranged in age from 21 to 27 years (M = 23.2, SD = 1.6). All subjects had completed at least a 4-year college degree before entering UPT. Subjects were informed that their performance on the BAT battery would not affect their continuation in UPT, would not be entered into their permanent service records, and would be used only for developing an improved USAF pilot candidate selection model.

Instrumentation

AFOQT. The AFOQT is a paper-and-pencil aptitude test battery used to select civilian or prior-service applicants for officer precommissioning training programs and to classify commissioned officers into aircrew job specialties (pilot versus navigator). The battery consists of 16 subtests that assess five ability domains: verbal, quantitative, spatial, aircrew interests/aptitude, and perceptual speed (Skinner & Ree, 1987). Fourteen of the AFOQT subtests are used to compute the Pilot and Navigator-Technical composite scores used in the operational selection of pilot candidates (US Air Force, 1983).

Basic Attributes Test (BAT). The BAT battery used in this study consisted of eight computerized tests that assessed individual differences in psychomotor coordination, information processing ability, personality, and attitudes. The types of scores generated include tracking error, response time, response accuracy, and response choice. Table 1 provides a brief summary of this battery. A more detailed description is provided by Carretta (1987, 1989).

Table 1. **Basic Attributes Test (BAT) Battery Summary**

Test name	Length (mins)	Attributes measured	Types of scores
Two-Hand Coordination (rotary pursuit)	10	Tracking and Time-Sharing Ability in Pursuit	Tracking error
Complex Coordination (stick and rudder)	10	Compensatory Tracking Involving Multiple Axes	Tracking error
Encoding Speed	20	Verbal Classification	Response time, response accuracy
Mental Rotation	25	Spatial Transformation and Classification	Response time, response accuracy
Item Recognition	20	Short-Term Memory, Storage, Search and Comparison	Response time, response accuracy
Time-Sharing	30	Higher-Order Tracking Ability, Learning Rate and Time-Sharing	Tracking difficulty, response time, dual-task performance
Self-Crediting Word Knowledge	10	Self-Assessment Ability, Self-Confidence	Response time, response accuracy, bet
Activities Interest Inventory	10	Survival Attitudes	Response time, number of high-risk choices

UPT performance criterion. UPT final training outcome was scored as a dichotomous variable, with graduates receiving a score of 1 and eliminees a score of 0.

Apparatus

The BAT apparatus consisted of a microcomputer and monitor built into a ruggedized chassis with a glare shield and side panels designed to minimize distractions. The subjects responded to the tests by manipulating (individually or in combination) a dual-axis joystick on the right side, a single-axis joystick on the left side, and a keypad in the center of the test unit. The keypad included keys labled 0 to 9, an ENABLE key in the center, and a bottom row with YES and NO keys, and two others for same/left (S/L) responses, and different/right (D/R) responses.

Procedure

Each subject was administered both the AFOQT and BAT prior to entry into UPT. Pilot candidates were commissioned through either AFROTC or OTS. Those from AFROTC were tested on the AFOQT prior to entering college or while an undergraduate. AFROTC pilot candidates were administered the BAT while attending a Flight Screening Program (FSP) in the summer following their junior year in college. For the OTS pilot candidates, the AFOQT was administered either just prior to or after completion of college, and the BAT was administered at the beginning of FSP.

The eight BAT tests used in this study were included in a longer battery that required about 3 1/2 hours to complete. After the test administrator initiated the battery, the test session was self-paced. Programmed breaks of 1 or 2 minutes were included between tests in order to reduce mental and physical fatigue.

All pilot candidates went through the same UPT program, which consisted of T-37 (initial jet trainer) and T-38 (advanced jet trainer) training. UPT graduates completed an average of 190 hours of flying. The final training outcome was determined at the end of the program.

Approach

To be useful as an adjunct to currently used pilot candidate selection procedures, the BAT performance measures must demonstrate incremental validity against training outcome when used in combination with operational instruments (i.e., AFOQT scores) and must demonstrate stability when cross-validated. For the cross-validations, the pilot candidates were divided randomly into two groups. The assignments were made such that the groups were similar in their UPT pass/fail rates. Table 2 provides a summary of the composition of each group. Pilot candidate selection models were developed independently for each group, using UPT final outcome as the training criterion.

Table 2. **UPT Outcome by Group**

	Group 1		Group 2	
	N	%	N	%
Graduates	285	66.3	300	65.9
Eliminees (total)	145	33.7	155	34.1
FTD	90	20.9	102	22.4
Academic	1	0.2	1	0.2
Medical	18	4.2	14	3.1
MOA	17	4.0	8	1.8
SIE	19	4.4	30	6.6

Note. FTD = Flying Training Deficiency.
MOA = Manifestation of Anxiety.
SIE = Self-Initiated Elimination.

For each half-sample, two approaches were used to develop "best-fitting" regression models for predicting UPT final outcome. One approach used a stepwise regression technique whereas the other simultaneously entered all 19 AFOQT/BAT scores (2 AFOQT composites and 17 BAT summary scores).

The stepwise inclusion method combined forward inclusion of the 19 AFOQT/BAT scores with deletion of scores that no longer contributed significantly to the regression solution. This stepwise method did not force any scores into the model initially. The probability for scores to enter and probability to leave the regression equation were set at .20 (Kim & Kohout, 1975). Regression weights from each half-sample were applied to the other half-sample to cross-validate the models and provide an estimate of shrinkage in the validity coefficients. The two half-samples were then combined to provide a best estimate of the regression weights for the pilot candidate selection models.

III. RESULTS

Prediction of UPT Final Outcome

Stepwise Regression Approach. Table 3 presents the results of regressing various predictor combinations on UPT final outcome. The objective was to identify the best combination of predictors to use in support of pilot selection decisions.

Table 3. **Prediction of UPT Final Outcome**

Method/Sample	N	N scores	UPT Pass rate	R	r
Stepwise					
Group1	430	9	.663	.290*	.211*
Group 2	455	9	.659	.315*	.202*
Combined	885	12	.661	.291*	
Simultaneous					
Group 1	430	19	.663	.306*	.227*
Group 2	455	19	.659	.326*	.242*
Combined	885	19	.661	.295*	

Note. The column labeled "R" indicates the multiple correlation of the model based on the regression weights for that group. The column labeled "r" indicates the correlation of the predicted outcome with actual outcome based on the regression weights from the other group (cross-validation).
*$p \leq .01$.

Results from the stepwise regression analyses suggested that individual differences on the AFOQT and BAT batteries were related strongly to final training outcome (Group 1: $R = .290$, $p \leq .01$; Group 2: $R = .315$, $p \leq .01$). Table 4 lists the test scores that were included in the final stepwise solution for each group. A comparison between the stepwise UPT models for the two half-samples indicated that they shared five common predictors from the AFOQT and BAT batteries. These included the AFOQT Pilot composite, average tracking difficulty from the Time-Sharing test and three response time scores (Mental Rotation, Item Recognition and Activities Interest Inventory). For both groups, subjects with good hand-eye coordination who made quick decisions were more likely to complete training successfully. When the regression weights from each half-sample were applied to the other half-sample to cross-validate the models, some reduction in the validity coefficients was observed. However, the cross-validated models were statistically significant (Group 1: $r = .211$, $p \leq .01$; Group 2: $r = .202$, $p \leq .01$).

A final stepwise regression model was developed using the entire sample (Group 1 and Group 2 combined). The 12-variable AFOQT/BAT model ($R = .291$, $p \leq .01$) included 11 of the 13 unique scores that contributed to one or both of the half-sample UPT models and added one test score that did not contribute to either half-sample (AFOQT Navigator-Technical composite).

Simultaneous Regression Approach. A best-fitting model that forced all 19 AFOQT and BAT scores into the regression equation was related significantly to UPT final outcome for both half-samples (Group 1: $R = .306$, $p \leq .01$; Group 2: $R = .326$, $p \leq .01$). Although the simultaneous regression solutions include redundant test scores that do not contribute significantly in the stepwise solutions, the simultaneous solutions generally appear more stable when cross-validated, demonstrating less shrinkage in the validity coefficients (Group 1: $r = .227$, $p \leq .01$; Group 2: $r = .242$, $p \leq .01$).

Table 4. **Comparison of Stepwise UPT Final Outcome Prediction Models by Group**

Test score	Group 1	Group 2	Combined
AFOQT			
Pilot	X	X	X
Navigator-Technical			X
Two-Hand Coordination			
X Axis Tracking Error	X		X
Complex Coordination			
X Axis Tracking Error		X	X
Y Axis Tracking Error			
Z Axis Tracking Error			
Encoding Speed			
Average Response Time	X		X
Mental Rotation			
Average Response Time	X	X	X
Percent Correct	X		
Item Recognition			
Average Response Time	X	X	X
Percent Correct			
Time-Sharing			
Average Tracking Difficulty	X	X	X
Average Response Time			
Self-Crediting Word Knowledge			
Average Response Time		X	X
Percent Correct	X		X
Bet			
Activities Interest Inventory			
Average Response Time	X	X	X
Number of High-Risk Choices		X	

Prediction of UPT Graduation/FTD Outcome

As shown in Table 2, over one-third of the UPT eliminations in this study were attributed to reasons other than poor flying training performance (Group 1: 55 of 145 eliminees = 37.9%, and Group 2: 53 of 155 eliminees = 34.2%, were for reasons other than flying training deficiency [non-FTD]). The AFOQT and BAT batteries are not designed to detect medical deficiencies, poor stress tolerance or lack of motivation. Therefore, a more appropriate estimate of the predictive validity of the AFOQT and BAT batteries for supporting pilot selection decisions would be made if only UPT graduates and FTD eliminees were included in the analyses.

Stepwise regression approach. The stepwise regression approach yielded significant multiple correlations with UPT/FTD outcome in both groups (Group 1: R = .327, $p \leq$ .01; Group 2: R = .365, $p \leq$ .01). Table 5 summarizes the regression results, and Table 6 lists the test scores that were included in the final stepwise solution for each group.

Table 5. **Prediction of UPT/FTD Outcome**

Method/Sample	N	N scores	UPT Pass rate	R	r
Stepwise					
Group1	375	8	.760	.327*	.257*
Group 2	402	9	.746	.365*	.220*
Combined	777	11	.753	.330*	
Simultaneous					
Group 1	375	19	.760	.344*	.249*
Group 2	402	19	.746	.380*	.260*
Combined	777	19	.753	.338*	

Note. The column labeled "R" indicates the multiple correlation of the model based on the regression weights for that group. The column labeled "r" indicates the correlation of the predicted outcome with actual outcome based on the regression weights from the other group (cross-validation).
*$p \leq$ .01.

When the regression weights from each half-sample were cross-validated with the other half-sample, the resulting validity coefficients were significant for both groups (Group 1: r = .257, $p \leq$ .01; Group 2: r = .220, $p \leq$ .01).

A final stepwise model was developed using the entire sample (Group 1 and Group 2 combined [n = 777]) to determine a "best estimate" of the regression equation. This model included 11 of the 13 test scores that contributed to one or both of the half-sample regression solutions (R = .330, $p \leq$ .01).

Simultaneous Regression Approach. As with the UPT final outcome analyses, a best-fitting model that forced all 19 AFOQT and BAT test scores into the regression equation was developed and cross-validated. This model was related significantly to UPT graduation/FTD outcome for both of the half-samples (Group 1: R = .344, $p \leq$ .01; Group 2: R = .380, $p \leq$ .01) and was cross-validated successfully (Group 1: r = .249, $p \leq$ .01; Group 2: r = .260, $p \leq$ .01). Although the stepwise and simultaneous regression solutions do not differ significantly in predictive validity during the model development phase, the simultaneous solutions generally appear more stable when cross-validated.

IV. DISCUSSION

Results from the AFOQT/BAT cross-validation regression analyses were consistent with those from Carretta (1989). For each half-sample, individual differences in psychomotor skills, information processing abilities, personality and attitudes helped to reduce uncertainty in making pilot candidate selection decisions. Further, results from the model development and cross-validation phases suggest that the selection models are robust. The significance of the cross-validation analyses is especially important because it indicates suitability for use in an operational setting.

Table 6. **Comparison of Stepwise UPT/FTD Outcome Prediction Models by Group**

Test score	Group 1	Group 2	Combined
AFOQT			
Pilot	X	X	X
Navigator-Technical			
Two-Hand Coordination			
X Axis Tracking Error	X		
Complex Coordination			
X Axis Tracking Error			
Y Axis Tracking Error			
Z Axis Tracking Error		X	X
Encoding Speed			
Average Response Time	X		X
Percent Correct			
Mental Rotation			
Average Response Time	X		
Percent Correct		X	X
Item Recognition			
Average Response Time	X	X	X
Percent Correct			
Time-Sharing			
Average Tracking Difficulty	X	X	X
Average Response Time		X	X
Self-Crediting Word Knowledge			
Average Response Time		X	X
Percent Correct	X		X
Bet			
Activities Interest Inventory			
Average Response Time	X	X	X
Number of High-Risk Choices		X	X

The validity estimates provided by the cross-validated selection models may seem low. It should be noted, however, that the magnitude of these correlations was limited by several factors. To begin with, there may have been some restriction in range on the abilities measured by the AFOQT and BAT batteries because these subjects had already been screened on the basis of their academic performance, aptitude test scores (i.e., AFOQT), and flying performance in a Flight Screening Program. Other factors that may have limited the magnitude of the correlations include (a) the dichotomous nature of the UPT outcome measures (UPT graduation versus elimination; UPT graduation versus FTD elimination) and (b) the proportion of UPT graduates in the sample (66.1% graduates). Results from the UPT/FTD analyses suggest that the validity of the AFOQT and BAT batteries against pilot training performance improves when pilot candidates who were eliminated for reasons other than FTD are removed from the analyses. A more sensitive training performance measure (e.g., flying grades, class standing) also may yield larger validity coefficients.

V. CONCLUSION

A combination of AFOQT and BAT performance scores demonstrated utility for supporting USAF pilot candidate selection decisions. The model development phase indicated substantial agreement between the stepwise selection models that were developed independently for the two groups. Although the simultaneous, forced-entry selection models did not improve the prediction of training outcome in the development phase beyond that provided by the stepwise approach, results show that the simultaneous solutions were more stable during the cross-validation phase (the simultaneous regression models showed less shrinkage in the validity coefficients when cross-validated). These models appear to be sufficiently robust to be used as adjuncts to operational USAF pilot candidate selection procedures.

REFERENCES

Bordelon, V.P., & Kantor, J.E. (1986). *Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies* (AFHRL-TR-86-4, AD-A170 353). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1987). *Basic Attributes Test (BAT) System: Development of an automated test battery for pilot selection* (AFHRL-TR-87-9, AD-A185 649). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1989). USAF pilot selection and classification systems. *Aviation Space and Environmental Medicine, 60, 1*, 46-49.

Kantor, J.E., & Carretta, T.R. (1988). Aircrew selection systems. *Aviation Space and Environmental Medicine, 59, 11* (Supplement), A32-A38.

Kim, J., & Kohout, F.J. (1975). Multiple regression analysis. In N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, & D.H. Bent, (Eds.). *Statistical Package for the Social Sciences* (2nd ed.). New York: McGraw-Hill.

McGrevy, D.F., & Valentine, L.D., Jr. (1974). *Validation of two aircrew psychomotor tests* (AFHRL-TR-74-4, AD-777 830). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Passey, G.E., & McLaurin, W.A. (1966). *Perceptual-psychomotor tests in aircrew selection: Historical review and advanced concepts* (PRL-TR-66-4,AD-636 606). Lackland AFB, TX: Personnel Research Laboratory.

Skinner, J., & Ree, M.J. (1987). Air Force Officer Qualifying Test (AFOQT): Item and factor analysis of Form O (AFHRL-TR-86-68, AD-A184 975). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

United States Air Force. (1983). *Application procedures for UPT, UPTH and UNT.* (Air Force Regulation 51-4). Washington, DC: Department of the Air Force.

* U. S. GOVERNMENT PRINTING OFFICE: 1990--761-051/20043