AL-TP-1991-0055

AD-A248 485

ARMSTRONG LABORATORY

PREDICTING PILOT TRAINING PERFORMANCE: DOES THE CRITERION MAKE A DIFFERENCE?

Thomas R. Carretta

DTIC ELECTE APR 0 8 1992 S B D

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

March 1992

Interim Technical Paper for Period 24 June 1991 - 21 October 1991

Approved for public release; distribution is unlimited.

92-08958

92 4 08 070

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

NOTICES

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 Office of Public Affairs 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.

Thomas R. Carretta

THOMAS R. CARRETTA
Project Scientist

William E. Alley

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Research Division

W. Alford

ROGER W. ALFORD, Lt Colonel, USAF
Chief, Manpower and Personnel Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

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

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1992

3. REPORT TYPE AND DATES COVERED
Interim 24 Jun 91 - 21 Oct 91

4. TITLE AND SUBTITLE
Predicting Pilot Training Performance:
Does the Criterion Make a Difference?

5. FUNDING NUMBERS
PE - 62205F
PR - 7719
TA - 18
WU - 45

6. AUTHOR(S)
Thomas R. Carretta

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armstrong Laboratory
Human Resources Directorate
Manpower and Personnel Research Division
Brooks Air Force Base, TX 78235-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
AL-TP-1991-0055

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)*

Traditionally, the utility of personnel attribute data for predicting U.S. Air Force pilot training performance has been evaluated against dichotomous training indicators (i.e., graduation or elimination, fighter or nonfighter aircraft recommendation). Several alternate Undergraduate Pilot Training (UPT) performance criteria based on flying performance data (i.e., daily flying grades, check flight grades, and academic grades) were evaluated to determine whether they could add to our understanding of the relationship between preselection personnel attribute data and UPT performance, beyond that provided by currently used dichotomous training performance indicators. UPT rankings were closely related to post-UPT follow-on training recommendations (better students were more likely to be recommended for fighter aircraft assignments). However, when the ranking algorithm was modified to include UPT eliminees, it demonstrated little utility in improving our understanding of the relationship between preselection personnel attribute data (i.e., test scores) and training performance beyond that provided by the UPT final outcome (graduation v. elimination) indicator.

14. SUBJECT TERMS
Undergraduate pilot training
Pilot candidate selection
Training performance
Personnel tests

15. NUMBER OF PAGES
20

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

CONTENTS

TABLES

PREFACE

This project was performed 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 William Glasscock, Sgt Steve Larsen, and Sgt Rob Long for their efforts in preparing the data files and programming the data analysis, and to Mr Gene Ligon and Ms Melinda Sanchez for administrative support. I also extend thanks to Maj Dave Perry, Dr Malcolm Ree, Dr Joseph L. Weeks, and Dr William E. Alley for their guidance and technical support during this project.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

PREDICTING PILOT TRAINING PERFORMANCE: DOES THE CRITERION MAKE A DIFFERENCE?

SUMMARY

The criteria used to represent United States Air Force pilot training performance typically have been dichotomous outcome indicators (graduation or elimination; fighter or non-fighter assignment). Although several valid predictors of training performance have been identified, it was felt that our understanding of the relationship between preselection personnel attribute data and training performance was limited by the dichotomous nature of the outcome indicators and by the disproportionate number of people in the outcome categories (i.e., the proportion of graduates in the Undergraduate Pilot Training [UPT] program is about 75%).

UPT rankings based on flying performance data (i.e., daily flying grades, check flight grades, and academic grades) were shown to be related closely to advanced training recommendations (fighter v. nonfighter aircraft). The data suggested that this ranking algorithm was a reasonable measure of pilot candidate quality as fighter aircraft assignments are considered more prestigious and demanding than nonfighter assignments.

When the ranking algorithm was modified to include UPT eliminees, however, it demonstrated little utility in adding to our understanding of the relationship between performance on selection instruments (i.e., test scores) and training performance. For pilot candidate selection purposes, the training criterion used to estimate the regression weights for the selection equation had little impact on the rankings of the applicants. These results were not surprising, however, as the dichotomous UPT final outcome indicator was strongly correlated with UPT performance as measured by the ranking algorithm.

INTRODUCTION

Since World War I, the United States (U.S.) military has used personnel tests to assess individual differences in attributes to make selection and classification decisions for pilot training applicants. These tests have included paper-and-pencil aptitude tests (e.g., measures of general intelligence, vocabulary, spatial ability, perception; see Skinner & Ree, 1987 for a description of the Air Force Officer Qualifying Test), and several apparatus based measures of perceptual and motor abilities (e.g., rotary pursuit, stick and rudder, compensatory tracking; see Imhoff and Levine, 1981 for a review of the literature). The United States Air Force (USAF) pilot research emphasis, largely, has been on the development and validation of sources of personnel attribute data to reduce training attrition and to capture policy decisions regarding specialized training suitability for bomber, fighter, tanker, or transport aircraft (Carretta, 1989).

The criteria used to represent pilot training performance typically have been dichotomous (i.e., graduation or elimination; fighter or nonfighter assignment). Although several valid predictors of flying training outcome have been identified, it was felt that our understanding of the relationship between these predictors and training performance was limited by the dichotomous nature of the outcome indicators (Cohen, 1983) and by the disproportionality in the outcome categories (Gradstein, 1986). The proportion of graduates in the Undergraduate Pilot Training (UPT) program typically is about 75%. Dichotomization of the training criteria resulted in reduction in the criterion variance accounted for by the predictors and reduction in statistical power (Cohen, 1983). A 75% graduation rate in pilot training would impose an upper limit of .734 on the point-biserial correlation between the predictors and a dichotomous final training outcome indicator (Gradstein, 1986).

The goals of this study were to (a) examine different procedures for generating training performance criteria that would reflect the relative quality of USAF pilot candidates (e.g., class rankings) based on flying performance scores and academic grades and to (b) evaluate the utility of these criteria for improving our understanding of the relationship between selection test scores and training performance. To be useful in a pilot candidate selection context (i.e., reduce attrition), a selection algorithm predicting an alternate training performance criterion (i.e., class ranking) must rank-order applicants in a more optimal manner than does the selection algorithm used to predict final training outcome (graduation or elimination).

METHOD

Subjects

The subjects used in this study were 755 USAF UPT students who were tested on both the Air Force Officer Qualifying Test (AFOQT; 696 Form O, 59 Forms M, N, or P) and Basic Attributes Test (BAT) batteries. All subjects had already been chosen for UPT, in part, on the basis of their AFOQT scores. The BAT battery was not part of the operational USAF pilot candidate selection procedure but is expected to become an operational selection instrument in 1992.

Subjects ranged in age from 21 to 31 years with an average of 24.7 years and were predominantly male (744 males, 11 females) and White (730 Whites, 25 non-Whites). 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 pilot candidate selection model. No subjects declined to participate.

Instrumentation

Air Force Officer Qualifying Test. The AFOQT is a paper-and-pencil multiple aptitude test battery used to select civilian or prior service applicants for officer precommissioning training programs and to classify commissioned officers into aircrew specialties (pilot v. navigator). The battery consists of 16 subtests that assess 5 ability domains: verbal, quantitative, spatial, perceptual speed, and aircrew interests/aptitude (Skinner & Ree, 1987). Fourteen of the 16 AFOQT subtests are used to compute the Pilot and Navigator-Technical composite scores used in the operational selection of pilot candidates (United States Air Force, 1983).

Basic Attributes Test. The BAT battery consisted of 8 computerized tests that assessed individual differences in psychomotor coordination (rotary pursuit, stick and rudder, compensatory tracking), information processing ability (reasoning, spatial transformation, short-term memory, perceptual speed), personality (self-confidence), and attitudes toward risk taking. The scores included tracking error/tracking difficulty, response time, response accuracy, and response choice. A more detailed description of the test battery, administration, and scoring procedures was provided by Carretta (1989).

Procedure

Prior to entry into UPT, each subject was administered both the AFOQT and BAT batteries. The AFOQT was administered prior to evaluation for an officer commissioning program (i.e., Reserve Officer Training Corps or Officer Training School). The BAT was administered at the beginning of a 2-week, light aircraft, flight screening program.

UPT is a 53-week program which consists of an academic Phase I concurrent with a T-37 Phase II (initial jet trainer, 21 weeks) and a T-38 Phase III (advanced jet trainer, 32 weeks).

UPT Performance Criteria

UPT final outcome. Final training outcome is typically scored as a dichotomous variable with graduates receiving a score of one and eliminees a score of zero. UPT graduates are evaluated for advanced training assignments (bomber, fighter, tanker, or transport aircraft) at the 43rd week of training by the training Wing Commander. Both final training outcome and advanced training assignment are determined, to a large degree, by academic grades, daily flying grades, and check flight grades.

Academic grades. Phase I (academic) indicators represented pilot candidates' performance on written tests of flying theory and procedures taken during UPT and were rated on a 4-point scale: (0) poor, (1) fair, (2) good, and (3) excellent. Academic Average (AA) reflects the number of points achieved on written tests as a ratio of the number of points possible, and may range from 0 to 100 (i.e., AA = [No. points achieved/No. points possible] × 100). AA is not calculated separately for T-37 and T-38 training.

Daily flying grades. These grades include instructor pilots' evaluations of a pilot candidate's flying performance on all flights other than check flights. Daily flying grades represented a weighted average of all flying procedures/maneuvers performed on a particular day and were rated: (0) poor, (1) fair, (2) good, and (3) excellent. Daily Flying Average (DFA) reflects the number of points achieved on all flights other than check flights as a ratio of the number of points possible, and may range from 0 to 100 (i.e., DFA = [No. points achieved/No. points possible] × 100). DFA is computed separately for Phase II and Phase III training.

Check flight grades. During UPT, a pilot candidate must pass a check flight in each of 10 courses of instruction--basic, contact, instrument, formation, and navigation flight maneuvers for both Phase II (T-37, basic jet trainer) and Phase III (T-38, advanced jet trainer). As with daily flying grades, check flight grades were a weighted average of ratings of flying procedures/maneuvers which may range from (0) poor to (3) excellent. Check Flight Average (CFA) reflects the number of points achieved on check flights as a ratio of the number of points possible (i.e., CFA = [No. points achieved/No. points possible] × 100). As with DFA, CFA is computed separately for Phase II and Phase III training.

Flying hours. The number of flying hours completed by each pilot candidate is recorded separately for Phase II (T-37) and Phase III (T-38) training. UPT graduates typically complete about 190 flying hours during the program.

Approach

To be useful for research purposes, a flying training criterion should (a) reflect the relative quality of the performance of all pilot candidates (both graduates and eliminees), (b) be based on overall performance rather than a specific flying maneuver, test score, or course of instruction, and (c) help to improve our understanding of the relationship between scores (i.e., test scores, biodata) and training performance beyond that provided by the dichotomous training criterion. To produce a stable performance indicator, the training criterion should incorporate as much training performance data as possible.

Criterion Development

United States Air Force Air Training Command (ATC) has used a UPT evaluation score based on UPT academic and flying grades for tracking and program evaluation purposes only (Corcoran, 1988). The evaluation score was a weighted average of Phase II (T-37) and Phase III (T-38) flying performance grades and Phase I (academic) grades. The score algorithms may be summarized as follows:

$$RNKIND = \frac{\text{Phase II PHA} + 2\ \text{Phase III PHA} + 0.5\ \text{AA}}{3.5} \quad (1)$$

where:

RNKIND = UPT Ranking Index

$$PHA = \text{Phase Average} = \frac{0.75\ (DFA) + CFA}{1.75} \quad (2)$$

where:

DFA = Daily Flying Average

CFA = Check Flight Average

AA = Academic Average

The weights for the RNKIND (Ranking Index) algorithm were arrived at through an "expert judgment" approach by experienced USAF instructor pilots. The RNKIND algorithm emphasizes the importance of check flight performance over daily flying performance and Phase III (T-38, advanced jet training) over Phase II (T-37, initial jet training). Relatively little weight is given to UPT academic performance (Phase I) in computing RNKIND.

The RNKIND score can range between 0 and 100 but in practice, UPT graduates generally score between 73 and 92.

As previously stated, the intended use of the RNKIND algorithm was as a program evaluation and tracking mechanism for UPT graduates. Trainees who receive a fighter recommendation for advanced training are generally perceived as superior to those who do not. If accurate, fighter-recommended trainees should receive higher RNKIND scores than nonfighter recommended trainees. To test this distribution relationship, all UPT graduates with a valid advanced training recommendation (488 out of 584 graduates) were rank-ordered from highest to lowest on this RNKIND score and divided into quintiles (i.e., 20% groups). Using a χ^2, a test against the uniform distribution was made. Rejection of the null hypothesis would indicate a relationship between the quintiles and RNKIND.

Several alternatives were considered for dealing with the UPT eliminees including: (a) removing them from the sample, (b) applying the RNKIND algorithm to eliminees without modification, (c) assigning all eliminees the same arbitrary score, and (d) using other available flying performance data (e.g., number of flying hours completed) to compute a ranking index score for eliminees. Removing the eliminees from the study was rejected because it would affect too many subjects (about 23%) and make it inappropriate to compare the ranking indices with the dichotomous training outcome indicator.

In addition to applying the RNKIND algorithm without modification to the UPT eliminees, 2 alternatives were considered. The first method arbitrarily assigned all eliminees a "ranking index" equal to 65 (RNK65). This value was chosen because it was below the lowest score for a UPT graduate, but not so low as to severely affect the variability of the score distribution. The second method computed "ranking index" scores for eliminees by taking into account the proportion of the training program completed (i.e., flying hours completed, RNKFLY). For UPT graduates, RNKFLY = RNKIND; for UPT eliminees:

$$\text{RNKFLY} = \frac{\textit{Total Flying Hours Completed}}{\textit{Maximum Flying Hours Completed by an Eliminee}} \times 70 \qquad (3)$$

The RNKFLY algorithm yields ranking index scores between 0 and 70 for UPT eliminees. An upper limit of 70 was used so that the highest scoring eliminee was below the lowest scoring graduate.

RESULTS

As shown in Table 1, training performance was strongly related to advanced training recommendation ($\chi^2[4] = 75.8$, $p \leq .01$). The proportion of fighter-recommended trainees decreases dramatically from the top to the bottom quintile. This result suggests that advanced training recommendations were made primarily on the basis of flying performance data, with an emphasis on Phase III (T-38) performance.

TABLE 1. FREQUENCY OF FIGHTER AND NONFIGHTER ADVANCED TRAINING RECOMMENDATION BY RNKIND

	Number of subjects receiving recommendation for		
RNKIND	Fighter	Nonfighter	% Fighter
Quintile			
1 (top 20%)	87	11	88.8
2	61	37	62.2
3	47	51	48.0
4	20	78	20.4
5 (bottom 20%)	15	81	15.6
Total	230	258	47.1

Note. Only 488 of the 584 UPT graduates had valid training recommendations.

Table 2 provides summary statistics of the score distributions for these ranking indices and for the dichotomous UPT final outcome measure.

TABLE 2. SCORE DISTRIBUTIONS FOR UPT PERFORMANCE CRITERIA (N = 755)

Criterion	Mean	SD	Minimum	Maximum	Skew	Kurtosis
UPT Final Outcome	0.77	0.17	0.0	1.0	----	----
RNKIND	70.1	24.9	0.0	91.8	-1.7	1.1
RNK65	78.4	7.7	65.0	91.8	-0.9	-0.6
RNKFLY	68.0	27.9	0.0	91.8	-1.5	0.6

RNKIND has been used only for tracking and program evaluation to evaluate the quality of UPT graduates. When the algorithm was applied to UPT eliminees, their RNKIND scores ranged between 0 and 74 because they received zeros for those phases they did not complete. Eliminees, therefore, demonstrated much more variability in UPT performance as measured by RNKIND (from 0 to 74) than did graduates (from 73 to 92). A few eliminees had higher RNKIND scores than the lowest ranking graduate. The RNKFLY algorithm yielded values between 0 and 70 for UPT eliminees.

To be useful in a pilot candidate selection context, the ranking index criteria should improve our understanding of the relationship between preselection factors and training performance and, as a result, allow us to make more optimal selection decisions (e.g., reduce attrition). Table 3 provides summary statistics of the distributions for the test scores used to predict the 4 UPT performance criteria (i.e., UPT final outcome, RNKIND, RNK65, and RNKFLY). It should be noted that many of the AFOQT and BAT score distributions are nonnormal and strongly skewed (i.e., BAT scores based on tracking performance or response time). Also, range restrictions occurred for the AFOQT Pilot and Navigator-Technical composites, as these pilot candidates had already been selected, in part, based on their AFOQT scores. Incidental range restriction occurred on the BAT variables as a function of their correlation with the AFOQT variables.

TABLE 3. SCORE DISTRIBUTIONS FOR BAT AND AFOQT BATTERIES

Score	Abbrv	Mean	Standard Deviation	Minimum	Maximum
2-Hand Coord					
Horiz Trk Err	PS2X1	10424.9	7489.9	2461.0	50653.0
Complex Coord					
Horiz Trk Err	PS2X2	9450.1	9901.9	228.0	72000.0
Vert Trk Err	PS2Y2	7914.9	12227.0	386.0	72000.0
Rudder Trk Err	PS2Z2	6492.7	6687.5	582.0	58155.0
Encoding Speed					
Avg RT (ms)	ENCRT	781.8	185.6	446.1	2157.0
% Correct	ENCPER	81.2	19.3	35.4	100.0
Mental Rotation					
Avg RT (ms)	MRTRT	928.6	520.1	88.3	7652.9
% Correct	MRTPER	90.9	9.4	45.8	100.0
Item Recognition					
Avg RT (ms)	ITMRT	842.0	226.6	430.9	2252.3
% Correct	ITMPER	95.0	4.3	62.5	100.0
Time-Sharing					
Avg RT (ms)	TMSRT	1172.5	241.1	664.8	3172.4
Trk Difficulty	TMSPER	260.0	37.2	112.8	335.6
Self-Cred Wd Know					
Avg RT (ms)	WKART	7604.8	1972.2	124.9	17009.0
% Correct	WKAPER	64.2	12.2	10.0	96.7
Avg Bet	WKABET	39.0	8.2	13.1	50.0
Act Interests Inv					
Avg RT (ms)	AIART	4442.8	1003.0	2120.0	8188.0
N H-Risk Choice	AIAHIR	49.0	12.2	12.0	80.0
Fly Experience	FLYEXP	6.4	4.4	1.0	20.0
AFOQT Plt Comp	PILOT	70.3	19.4	12.0	99.0
AFOQT Nav-Tec Comp	NAV	67.2	21.2	8.0	99.0

TABLE 4. CORRELATIONS AMONG UPT PERFORMANCE CRITERIA AND TEST SCORES

	Score	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	P/F	1.00																							
2	RNKIND	.91	1.00																						
3	RNK65	.95	.89	1.00																					
4	RNKFLY	.95	.98	.93	1.00																				
5	PS2X1	-.16	-.16	-.12	-.14	1.00																			
6	PS2X2	-.12	-.12	-.15	-.13	-.02	1.00																		
7	PS2Y2	-.12	-.13	-.14	-.13	.18	.45	1.00																	
8	PS2Z2	-.15	-.15	-.18	-.16	.04	.63	.37	1.00																
9	ENCRT	-.01	.02	.00	.01	.00	.16	.05	.10	1.00															
10	ENCPER	-.15	-.16	-.14	-.15	.42	-.10	.03	-.06	-.17	1.00														
11	MRTRT	-.06	-.05	-.07	-.05	.16	.03	.07	.06	.29	-.20	1.00													
12	MRTPER	.01	.03	.03	.03	-.02	-.10	-.10	-.13	-.13	.02	-.21	1.00												
13	ITMRT	-.12	-.08	-.11	-.08	.20	.09	.09	.08	.47	.13	.32	-.10	1.00											
14	ITMPER	.01	.03	.03	.03	.00	-.08	-.07	-.08	.12	.06	.04	.25	.08	1.00										
15	TMSRT	-.14	-.15	-.13	-.13	.38	.00	.07	.06	.14	.29	.15	-.05	.27	.05	1.00									
16	TMSDIF	.12	.11	.11	.11	-.18	-.13	-.17	-.22	-.15	.08	-.12	.05	-.17	.00	-.21	1.00								
17	WKART	-.03	-.03	-.02	-.02	.22	-.02	-.01	-.01	.12	.26	.11	-.03	.13	.03	.31	.04	1.00							
18	WKAPER	-.04	-.04	-.05	-.04	.08	-.04	-.04	.00	-.07	.21	-.10	.12	.03	.10	.01	.10	.11	1.00						
19	WKABET	.03	.02	.02	.03	-.01	-.04	-.03	-.03	-.06	-.02	-.09	.05	-.06	.05	-.10	.02	-.27	.32	1.00					
20	AIART	-.05	-.03	-.03	-.02	.05	-.02	-.02	-.03	.10	.13	.08	.06	.07	.12	.25	.06	.45	-.07	-.19	1.00				
21	AIAHIR	-.10	-.10	-.12	-.10	.02	-.04	-.07	-.02	-.10	.31	-.12	.00	.00	-.03	.07	.15	.20	.21	-.02	.04	1.00			
22	FLYEXP	.13	.15	.20	.16	-.08	.08	.02	.09	.19	-.35	.13	-.02	.10	-.05	-.01	-.12	-.04	-.17	-.03	.03	-.20	1.00		
23	PILOT	.12	.13	.18	.13	-.07	-.17	-.14	-.14	-.25	.11	-.23	.23	-.14	.01	-.18	.11	-.15	.18	.15	-.03	.06	.00	1.00	
24	NAV	.08	.08	.13	.09	-.05	-.14	-.11	-.13	-.29	.16	-.26	.25	-.15	.07	-.19	.13	-.17	.19	.18	-.06	.05	-.16	.87	1.0

Note: The first 4 scores are UPT final outcome and the 3 ranking indices. Scores 5 through 24 are BAT and AFOQT summary scores. See Table 3 for an explanation of the BAT and AFOQT score acronyms.

DISCUSSION

The results showed that the ranking of candidates was nearly identical for equations based on all the criteria. For pilot training candidates, the criterion did not make a difference as to who would have been selected.

The correlations of the expected scores was predictable by the magnitude of the correlations between the dichotomous UPT final outcome indicator and the UPT performance scores based on the ranking algorithm (r between .91 and .95).

Given the strength of agreement between the training criteria and in the pilot candidates' rankings on expected scores for the four UPT criteria, use of a training criterion based on flying performance data (i.e., flying grades) would not necessarily have resulted in a lower attrition rate than if the dichotomous UPT final outcome criterion was used.

CONCLUSION

UPT rankings generated from a training evaluation algorithm were shown to be related closely to advanced training recommendations (fighter v. nonfighter aircraft). This relationship suggests that the ranking algorithm is a reasonable indicator of pilot candidate quality, as fighter aircraft assignments are considered more prestigious than nonfighter assignments.

When the ranking algorithm was modified to include UPT eliminees however, it demonstrated little utility in adding to our understanding of the relationship between preselection personnel test scores and training performance. For pilot candidate selection, the training criterion used to estimate the regression weights for the selection equation had little impact on the ranking of the applicants once the predictors were held constant.

REFERENCES

Carretta, T.R. (1989). USAF pilot selection and classification systems. *Aviation Space and Environmental Medicine, 60*(1), 44-49.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurements, 7*(3), 249-253.

Corcoran, B.J. (telephone interview, 5 Aug 1988). Description of USAF Air Training Command Class Ranking Algorithm for UPT Students.

Gradstein, M. (1986). Maximal correlation between normal and dichotomous variables. *Journal of Educational Statistics, 11*(4), 259-261.

Imhoff, D.L., & Levine, J.M. (1981). *Perceptual-motor and cognitive performance task battery for pilot selection* (AFHRL-TR-87-20, AD-A094 317). Brooks AFB, TX.

Skinner, J., & Ree, M.J. (1987). *Air Force Officer Qualifying Test (AFOQT): Item and factor analyses of Form O* (AFHRL-TR-86-68, AD-A184 975). Brooks AFB, TX.

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