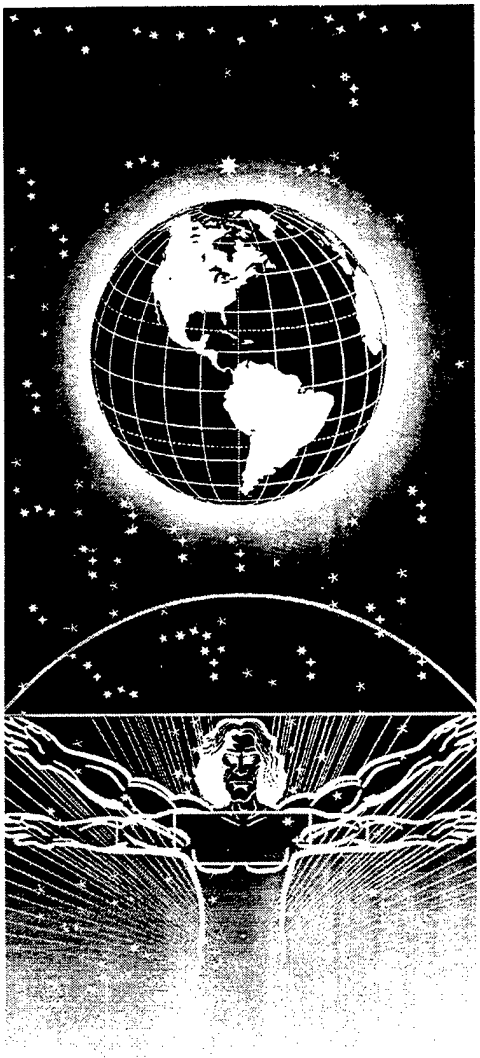
AL/HR-TP-1997-0040



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

BASIC ATTRIBUTES TEST
RETEST PERFORMANCE

Thomas R. Carretta
Warren E. Zelenski

AIRCREW PERFORMANCE BRANCH
AIRCREW TRAINING RESEARCH DIVISION

Malcolm James Ree

COGNITION AND PERFORMANCE DIVISION

HUMAN RESOURCES DIRECTORATE
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352

December 1997

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
ARMSTRONG LABORATORY
HUMAN RESOURCES DIRECTORATE
AIRCREW TRAINING RESEARCH DIVISION
6001 South Power Road, Building 558
Mesa AZ 85206-0904

19980325 066

NOTICES

Publication of this paper does not constitute approval or disapproval of the ideas or findings. It is published in the interest of scientific and technical information (STINFO) exchange.

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
Project Scientist

DEE H. ANDREWS
Technical Director

LYNN A. CARROLL, Colonel, USAF
Chief, Aircrew Training Research Division

Please notify AL/HRPP, 7909 Lindbergh Drive, Brooks AFB, TX 78235-532, if your address changes, or if you no longer want to receive our technical reports. You may write or call the STINFO Office at DSN 240-3877, or commercial (210) 536-3877.

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 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	3. REPORT TYPE AND DATES COVERED
	December 1997	Interim - January 1994 to August 1996

4. TITLE AND SUBTITLE

Basic Attributes Test Retest Performance

5. FUNDING NUMBERS

PE – 62205F
PR – 1123
TA – B1
WU – 01

6. AUTHOR(S)

Thomas R. Carretta
Warren E. Zelenski
Malcolm James Ree

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Aircrew Training Research Division
Aircrew Performance Branch
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352

Armstrong Laboratory
Cognition and Performance Division
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352

8. PERFORMING ORGANIZATION REPORT NUMBER

AL/HR-TP-1997-0040

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

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Thomas R. Carretta, (210) 536-3956.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Basic Attributes Test (BAT) is a component of a US Air Force pilot selection composite known as the Pilot Candidate Selection Method (PCSM). When PCSM was operationally implemented in 1993, no retests were permitted on the BAT. A study was conducted to examine retest reliability and mean score change for the BAT. Four hundred seventy-seven (477) college students completed the BAT and retested at one of three intervals: 2 weeks, 3 months, or 6 months. Several important findings were observed. First, BAT scores demonstrated acceptable reliability. Second, about 70% of the students exhibited score improvement on retest, regardless of length of retest interval. Those who performed poorly on the first test generally exhibited larger improvements than those who performed well on the first test. Third, practice effects diminished as the length of the retest interval increased. For a six-month retest interval, it is expected that PCSM scores would increase on average by about six percentile points. These results suggest that BAT retests should be permitted no less that six months after initial testing.

14. SUBJECT TERMS

Basic Attributes Test; BAT; PCSM, Pilot Candidate Selection Method; Pilot selection; Practice effects; Reliability; Retest

15. NUMBER OF PAGES

17

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	Unlimited

NSN 7540-01-280-5500

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

CONTENTS

TABLES

FIGURES

Preceding Page Blank

PREFACE

This effort was conducted under Work Unit 1123B101 (formerly 1123A101), Pilot Selection and Classification Support, which is dedicated to research into the selection and classification of US Air Force aircrew personnel. The laboratory work unit monitor is Dr Frederick M. Siem.

We thank Ms C. Patrick for her assistance in collecting the data.

Address correspondence and requests for reprints to the first author at AL/HRAA, 7909 Lindbergh Drive, Brooks Air Force Base, TX 78235-5352. Send e-mail to CARRETTA@ALHRM.BROOKS.AF.MIL.

BASIC ATTRIBUTES TEST RETEST PERFORMANCE

SUMMARY

The Basic Attributes Test (BAT) is a component of a US Air Force pilot selection composite known as the Pilot Candidate Selection Method (PCSM). When PCSM was operationally implemented in 1993, no retests were permitted on the BAT. A study was conducted to examine retest reliability and mean score change for the BAT. Four hundred and seventy-seven (477) college students completed the BAT and retested at one of three intervals: 2 weeks, 3 months, or 6 months. Several important findings were observed. First, BAT scores demonstrated acceptable reliability. Second, about 70% of the students exhibited score improvement on retest, regardless of length of retest interval. Those who performed poorly on the first test generally exhibited larger improvements than those who performed well on the first test. Third, practice effects diminished as the length of the retest interval increased. For a six-month retest interval, it was expected that PCSM scores would increase on average by about 6 percentile points. These results suggest that BAT retests could be permitted no less than six months after initial testing.

INTRODUCTION

The Basic Attributes Test (BAT) is a component of a US Air Force pilot candidate selection composite known as the Pilot Candidate Selection Method or PCSM (Carretta, 1992a; Carretta & Ree, 1994). Current USAF policy allows only one test on the BAT (see Air Force Instruction 36-2605, 17 June 94). When operationally implemented for pilot selection in 1993, BAT retest performance had not been adequately addressed. Carretta (1992b) examined retest performance (means, reliability) for a pre-operational form of the BAT, but the retest interval was only one day. Results indicated that about 70% of the participants showed score improvement in performance on retest. There were moderate to large mean score improvements, depending on test content. Test-retest reliability was acceptable. Although the study was informative, it is not likely the US Air Force would permit a retest after such a short interval with the possible exception of a test system computer failure during a test session. A more likely scenario would allow pilot applicants to retest on BAT after an interval of several weeks or months as is done with other selection instruments used by the US military (e.g., Armed Services Vocational Aptitude Battery or ASVAB, Air Force Officer Qualifying Test or AFOQT). For instance, Air Force applicants are permitted to retest on the AFOQT after an interval of at least six months.

The objective of this study was to examine mean score performance and reliability on the BAT in the event of a retest. These data could be used to inform policy makers regarding expected changes to mean test scores and rank-ordering of retesters if retests were allowed on the BAT.

METHOD

Participants

Four hundred and seventy-seven (477) Arizona State University (ASU) students participated in this study as paid volunteers. There were slightly more males than females in the participant sample (56% male, 44% female). Participants were informed that the study involved the evaluation of an operational US Air Force pilot selection test. Students who were enrolled in Air Force training programs or expressed a desire to someday enter the Air Force were not permitted to participate so that their scores of record would reflect current performance. Participants were told that their performance on the BAT would be entered into their permanent service records in the event they later decided to enter the US Air Force and apply for pilot training.

Measures

The BAT is a computer-based test battery used by the US Air Force for pilot selection. BAT scores are combined in a weighted equation with the AFOQT pilot composite and a measure of flying experience to produce a pilot aptitude index known as the Pilot Candidate Selection Method or PCSM (Carretta, 1992a). The short-term retest reliability of the BAT has been investigated (Carretta, 1992b) as has its validity for pilot training (Carretta, 1992a; Carretta & Ree, 1994), factor structure (Ree & Carretta, 1994), and group differences in performance (Carretta, 1997).

The operational BAT has five tests including Two-Hand Coordination (psychomotor), Complex Coordination (psychomotor), Item Recognition (short-term memory), Time Sharing (psychomotor), and Activities Interest Inventory (attitudes). The test apparatus consists of a 386-based computer and monitor built into a testing carrel. Participants respond to the test stimuli by manipulating individually, or in combination, a dual-axis, right-hand control stick; a single-axis left-hand control stick; and a specialized response keypad. See Carretta (1992b) for a more detailed description of the BAT tests and hardware.

Procedure

Each participant completed the BAT and was randomly assigned to one of three retest intervals: 2 weeks (n = 192), 3 months (n = 167), or 6 months (n = 118). Each participant retested on BAT at the completion of one of the retest intervals. No practice was permitted between the first and second test.

Analyses

Analyses consisted of examination of mean score changes on retest and the correlation between the first and second test scores.

Mean scores. Differences between first and second administration means were expressed in standard deviation units or *d* (i.e., $[\bar{X}_1 - \bar{X}_2]/ S_D$). The standard deviation for *d* was defined as the within-group standard deviation calculated from the weighted average of the square root of the variances for the scores being compared (e.g., first versus second test for the 2-week interval group). Frequently, *d* values are used as an estimate of effect size. It should be noted that *d* values of .20 are commonly considered "small," .50 "medium," and .80 "large" (Cohen, 1988). In addition to the computation of *d*, one-tailed, paired-samples t-tests were performed to examine whether performance *improved* on retest. A .05 Type I error rate was used for the t-tests.

Note that improvements in tracking error scores (Two-Hand Coordination and Complex Coordination) and response time (Item Recognition, Time Staring, and Activities Interest Inventory) will result in positive values for the *d* and t-tests. Improvements in the psychomotor composite, tracking difficulty (Time Sharing), and percentage scores (Item Recognition, Activities Interest Inventory) result in negative values for *d* and t-tests.

Test-retest correlations. Correlations between first and second test scores indicate the degree to which the rank order of participants on the first test change after retesting (i.e., is the ranking on the first test the same as the ranking on the second test?). Test-retest correlations also provide an estimate of reliability.

RESULTS

Results varied by test. See Tables 1 through 3 for a detailed summary of the mean score analyses and test-retest correlations for the 2-week, 3-month, and 6-month retest groups.

Psychomotor Composite

Two-Hand Coordination and Complex Coordination scores contribute to a BAT psychomotor composite. Performance significantly improved on retests for both of these tests. The amount of improvement declined as the retest interval increased. For the psychomotor composite, the amount of improvement on retest was d = .48, .33, and .25 for the 2-week, 3-month, and 6-month groups respectively.

The correlations between first and second test scores indicated high agreement in rank order between first and second tests and acceptable test-retest reliability. For the psychomotor composite, the correlations were .800, .801, and .775 for the 2-week, 3-month, and 6-month groups.

Item Recognition

No significant mean score improvement was observed for any of the retest groups. The test-retest correlations also were acceptable. The correlations for the percent correct scores were a little low indicating change in rank order, but are probably due to the high accuracy rate on this test (about 95%).

Table 1.
Basic Attributes Test Retest Scores (2-Week Retest Group, N = 192)

Test Score	First Test Average	Second Test Average	S_D	*d*	t	r_{12}
Two-Hand Coordination						
Horizontal Error	7730.34	5944.08	3078.04	0.58	8.02**	.677
Vertical Error	9222.31	6595.25	3043.40	0.86	11.93**	.670
Complex Coordination						
Horizontal Error	32878.63	26502.06	12258.68	0.52	7.19**	.800
Vertical Error	28594.98	23091.39	13227.08	0.42	5.75**	.760
Rudder Error	23229.19	18455.39	14117.45	0.34	4.68**	.720
Psychomotor Composite	-1.1365	-0.6941	0.928	-0.48	-6.59**	.800
Item Recognition						
Response Time	808.25	847.44	210.31	-0.19	-2.58	.744
% Correct	94.75	94.44	7.10	0.04	0.60	.518
Time Sharing						
Response Time	922.96	902.31	127.97	0.16	2.23*	.727
Tracking Difficulty	195.21	203.36	28.28	-0.29	-3.98**	.808
Activities Interest Inv.						
Response Time	3699.78	2894.81	801.81	1.00	13.88**	.775
% Choices	52.94	53.48	7.39	-0.07	-1.01	.860

Notes. All t-tests were one-tailed. r_{12} is the correlation between the first and second test score. S_D is the standard deviation of the difference between the first and second test administrations.
*p < .05; **p < .01

Time Sharing

The amount of mean score improvement for this psychomotor test was much less than found for the other psychomotor tests. As with Two-Hand Coordination and Complex Coordination, the amount of score improvement decreased as the test-retest interval increased. The mean score changes in the 6-month interval were not statistically significant.

The test-retest correlations declined for the response time score as the retest interval increased (.727, .625, and .474). However, the test-retest correlations for the tracking difficulty score were fairly stable (.808, .806, and .767) indicating preservation of rank ordering on retest.

Activities Interest Inventory

This attitude scale produces a response time score and a percent of "correct" choices. The response time score provides an index of decisiveness/compulsiveness. There was a large test-retest effect for the response time for each retest interval (d = 1.00, 0.71, and 0.64). No significant improvement was observed for the percent score.

The test-retest correlations were acceptable for both the response time (.775, .655, and .771) and percent of "correct" choices (.860, .871, and .856) scores.

Table 2.
Basic Attributes Test Retest Scores (3-Month Retest Group, N = 167)

Test Score	First Test Average	Second Test Average	S_D	d	t	r_{12}
Two-Hand Coordination						
Horizontal Error	7107.76	5697.45	3156.13	0.44	5.76**	.650
Vertical Error	8461.93	6364.95	2623.72	0.80	10.30**	.717
Complex Coordination						
Horizontal Error	29018.59	23495.73	10979.82	0.50	6.48**	.827
Vertical Error	25639.14	21293.44	12298.14	0.35	4.55**	.767
Rudder Error	18425.19	15055.73	13824.64	0.24	3.14**	.668
Psychomotor Composite	-0.7981	-0.5041	0.882	-0.33	-4.30**	.801
Item Recognition						
Response Time	774.25	752.32	188.03	0.12	1.50	.792
% Correct	95.07	94.39	5.06	0.13	1.73	.778
Time Sharing						
Response Time	912.01	881.55	163.90	0.19	2.39**	.625
Tracking Difficulty	207.72	210.69	26.48	-0.11	-1.45	.806
Activities Interest Inv.						
Response Time	3402.05	2857.19	764.81	0.71	9.18**	.655
% Choices	54.85	56.13	7.30	-0.18	-2.26	.871

Note. All t-tests were one-tailed. r_{12} is the correlation between the first and second test score. S_D is the standard deviation of the difference between the first and second test administrations.
*$p < .05$; **$p < .01$

Table 3.
Basic Attributes Test Retest Scores (6-Month Retest Group, N = 118)

Test Score	First Test Average	Second Test Average	S_D	*d*	t	r_{12}
Two-Hand Coordination						
Horizontal Error	7364.55	6216.77	2715.45	0.42	4.57**	.823
Vertical Error	8551.51	6755.17	2862.72	0.63	6.79**	.685
Complex Coordination						
Horizontal Error	31131.60	25887.16	12627.19	0.42	4.49**	.749
Vertical Error	26948.38	22713.47	12789.29	0.33	3.58**	.768
Rudder Error	19685.72	17524.89	12264.88	0.18	1.91*	.735
Psychomotor Composite	-0.9067	-0.6607	0.988	0.25	-2.69**	.775
Item Recognition						
Response Time	828.55	794.15	235.99	0.15	1.58	.758
% Correct	95.22	93.93	5.54	0.23	2.52	.525
Time Sharing						
Response Time	932.79	902.84	321.61	0.09	1.01	.474
Tracking Difficulty	203.60	203.45	30.18	0.01	0.05	.767
Activities Interest Inv.						
Response Time	3459.99	3001.68	720.65	0.64	6.88**	.771
% Choices	52.74	54.29	7.90	-0.20	-2.12	.856

Note. All t-tests were one-tailed. r_{12} is the correlation between the first and second test score. S_D is the standard deviation of the difference between the first and second test administrations.
*$p < .05$; **$p < .01$

BAT Composite and PCSM Percentile

A BAT composite score was computed for each set of test-retest data using the weights for BAT scores from the PCSM equation. Figure 1 shows BAT raw score composite retest change for the three retest groups. As expected (Stanley, 1971, p. 379), the greatest amount of score improvement is found for those with the lowest scores on the first test. Amount of retest gain decreased as the participant's first test score increased. Therefore, it appears that those who have the most to gain on a retest (i.e., those with low scores on a first test) do in fact show the greatest improvement.

Figure 2 shows BAT raw score change based on initial score quartile. It shows that as the length of the retest interval increased, the amount of BAT score improvement decreased.

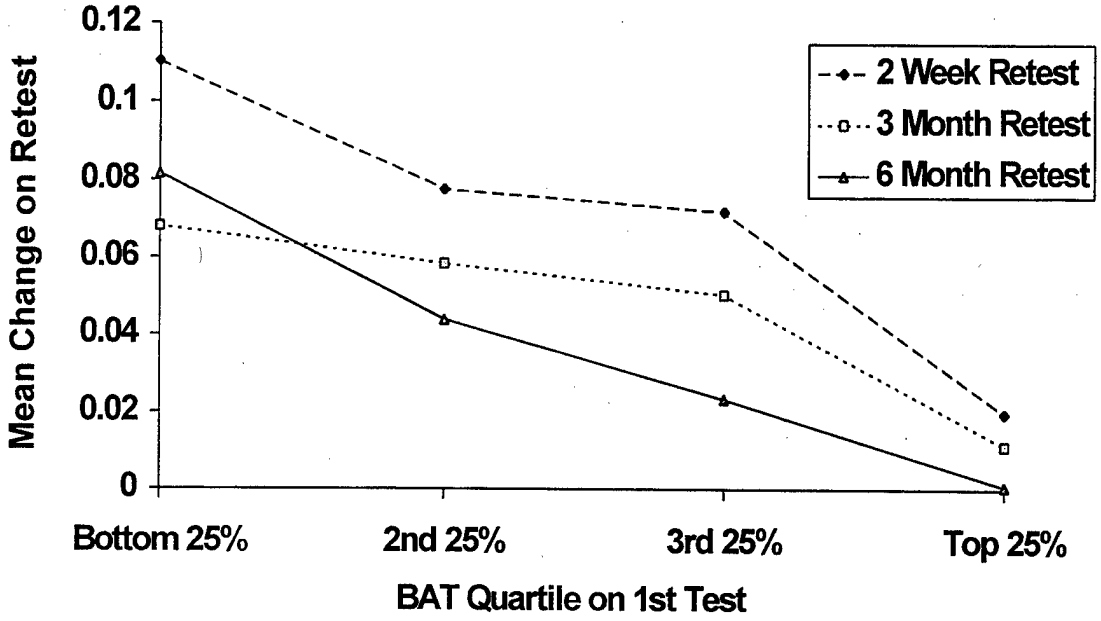


Figure 1. BAT Raw Score Composite Retest Change (by Retest Group)

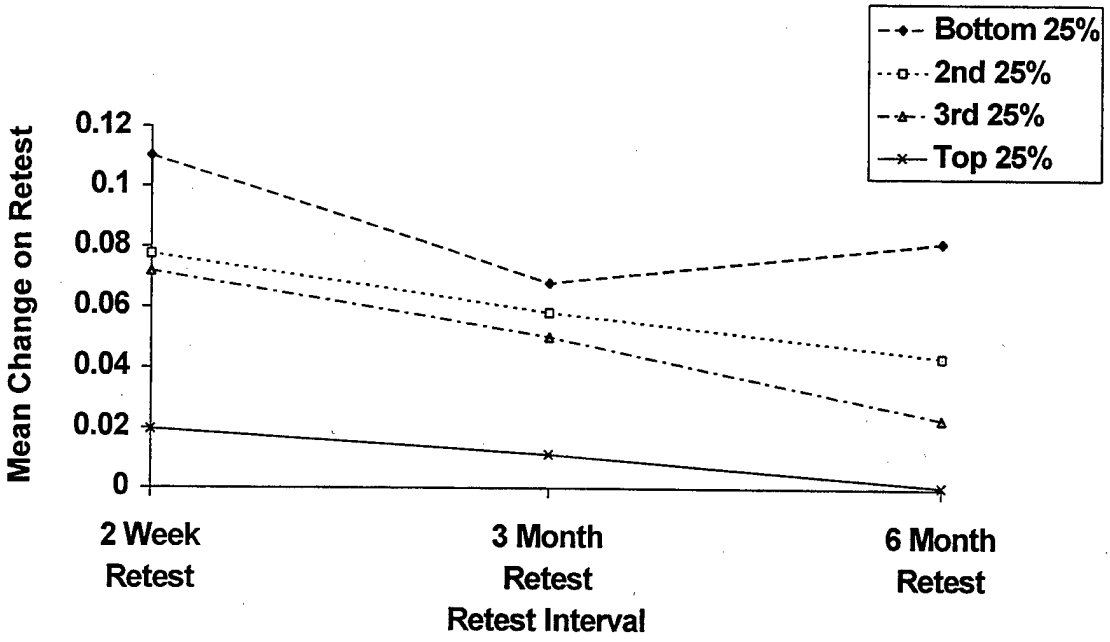


Figure 2. BAT Raw Score Composite Retest Change (by First Test Quartile)

Changes in BAT raw scores are manifested as changes in PCSM percentile scores. The conversion from raw score to percentile score is not linear. BAT raw score changes produce the greatest changes to PCSM percentiles in the 30 to 70 range. PCSM percentile scores below 10 or above 90 are relatively insensitive to changes in raw score. PCSM percentile score changes expressed in the following paragraphs assume an initial score in the 30 to 70 range and may, therefore, overestimate the impact of BAT raw score changes on PCSM percentile scores outside of this range.

ASU Students vs. US Air Force Pilot Applicants

ASU participants matched US Air Force pilot applicants (specifically, Air Force Reserve Officer Training Corps cadets) in terms of age and education. It should be noted that, as a group, the 5,254 Air Force pilot applicants who operationally tested on BAT as of the time of this study, outperformed the ASU college students. More than half of the Air Force pilot applicants achieved BAT scores that equaled or exceeded those of the top 25% of ASU students. Figure 3 shows the distribution of USAF pilot applicant scores relative to ASU quartiles.

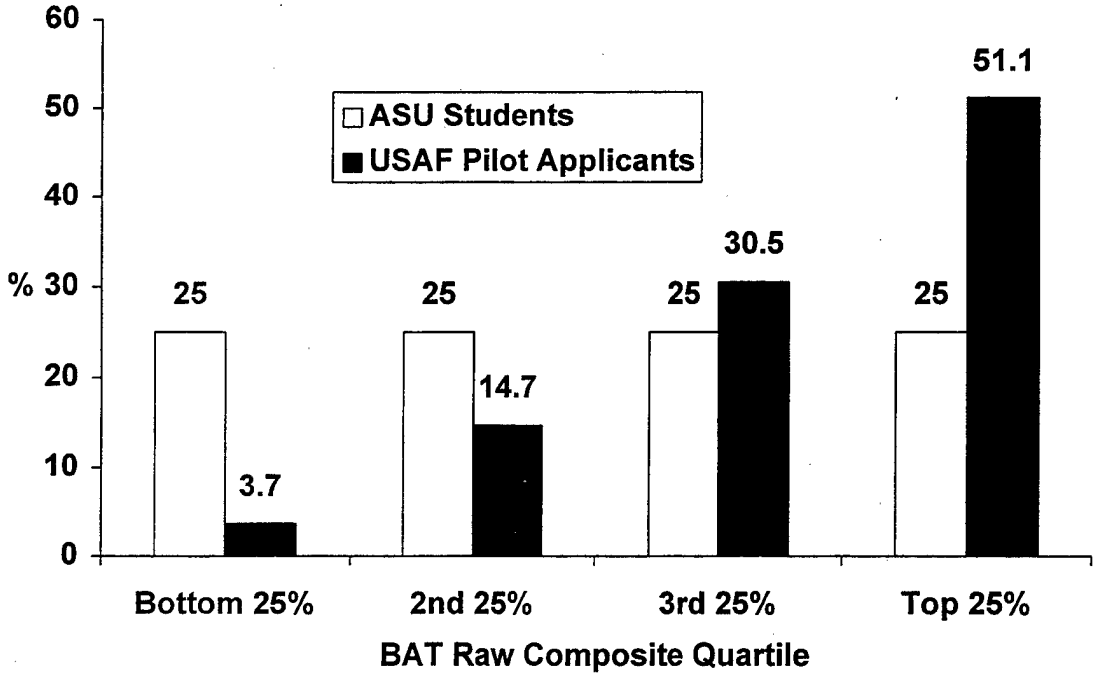


Figure 3. Proportion of US Air Force Pilot Applicants in ASU BAT Composite Quartiles

The ASU sample was weighted to produce a sample with initial BAT scores representative of US Air Force pilot applicants. BAT score changes were then computed for each of these retest intervals and expressed in terms of PCSM percentiles. The average expected change in PCSM percentile and the range of score changes defining the middle 50% of the sample are shown in Table 4.

Half of the individuals in the weighted ASU sample (i.e., the middle 50%) would experience score changes between -4 and +20 percentile points. The average PCSM change for this sample at six months is +6 percentile points.

Table 4.
Expected PCSM Percentile Change for US Air Force Pilot Applicants

Retest Interval	Average PCSM Change	Range of Middle 50%
2 Weeks	+13	0 to +33
3 Months	+10	-1 to +25
6 Months	+6	-4 to +20

DISCUSSION

The Air Force has longstanding retest policies for other personnel selection tests. Officer and aircrew training applicants may retest on the AFOQT after at least a six-month interval. Carretta and Ree (1997) examined performance for a group of over 40,000 officer applicants who voluntarily retested on AFOQT. Analyses consisted of examination of mean score changes on retest and the correlation between the first and second test scores for all retesters.

Large mean score improvements were found for each of the AFOQT composites for those who tested twice (Verbal, .79 *d* or 10 percentile points; Quantitative, .57 *d* or 8 percentile points; Academic Aptitude, .84 *d* or 10 percentile points; Pilot, .93 *d* or 14 percentile points; and Navigator-Technical, .91 *d* or 12 percentile points). Even larger improvements were observed for applicants who tested three or four times on the AFOQT. Therefore, retesters tend to improve their rank relative to those who choose not to retest.

Test-retest correlations were high for the AFOQT composites indicating that the rank order of the retesters (compared to other retesters) does not change as a result of a retest. For those who tested twice, the correlations were: Verbal, .885; Quantitative, .842; Academic Aptitude, .886; Pilot, .825; and Navigator-Technical, .866. Test-retest correlations between first and last tests declined for those who tested three or four times on the AFOQT.

In light of the results for the AFOQT, the amount of improvement observed on BAT scores is small. However, it should be noted that the participants in the AFOQT study were highly motivated to improve their performance as they were applying for officer commissioning or aircrew training and on average had scored below their expectations. Those in the ASU BAT sample, on the other hand, had no real incentive to try to improve their performance on a retest. It is speculated that greater improvement would have occurred if the sample had consisted of Air Force pilot applicants, especially if they voluntarily retested.

The BAT test-retest correlations suggested that the relative order of the retesters stayed the same after retesting. However, as with the AFOQT, retesters did tend to improve their mean score on retesting. Therefore, their scores will improve relative to those who choose not to retest.

CONCLUSIONS AND RECOMMENDATIONS

Several important findings were observed. First, BAT scores demonstrated acceptable retest reliability. That is, the rank-order of participants on BAT "first test" scores was about the same after retesting. Most BAT test-retest reliabilities were nearly as high as those for the AFOQT.

Second, mean score changes on some BAT tests (Two-Hand Coordination, Complex Coordination, Activities Interest Inventory) were comparable to results with the AFOQT composites. The BAT Item Recognition and Time Sharing tests showed relatively small mean score changes compared to the AFOQT.

Third, approximately 70% of the participants exhibited score improvement on the BAT composite on retest, regardless of length of retest interval. As expected, the amount of score improvement varied as a function of "first test" score. Those who performed poorly on the first test generally exhibited larger improvements than those who performed well on the first test. About 30% (mostly high scorers) showed no improvement or even a decrement on the second test. Therefore, we felt that it would be inappropriate to suggest a score adjustment for the second BAT test, especially when the participants were not motivated by qualification requirements.

Fourth, practice effects diminished as the length of the retest interval increased. For a 6-month retest interval, it is expected that PCSM percentile scores would increase on average by about 6 points. Fifty percent of the 6-month retesters would change between -4 and +20 PCSM percentiles. Twenty-five percent of retesters would improve their PCSM score by more than 20 percentiles. It should be noted, however, that BAT mean score changes for the 6-month retest group were smaller than typically found for another US Air Force pilot selection test, the AFOQT.

Based on these results, BAT retest policy could be changed to be consistent with current AFOQT retest policy. That is, a retest could be allowed after at least a 6-month test-retest interval. As with the AFOQT, the more recent score would be reported to selection boards and the retest scores would not be adjusted in any way. If the Air Force operationally implements BAT retesting, additional studies will need to be done to evaluate the validity of first versus later BAT scores.

REFERENCES

Carretta, T. R. (1992a). Recent developments in U. S. Air Force pilot candidate selection and classification. *Aviation, Space, and Environmental Medicine, 63,* 1112-1114.

Carretta, T. R. (1992b). Short term retest reliability of an experimental U. S. Air Force pilot candidate selection test battery. *The International Journal of Aviation Psychology, 2,* 161-173.

Carretta, T. R. (1997). Male-female performance on U. S. Air Force pilot selection tests. *Aviation, Space, and Environmental Medicine, 68,* 818-823.

Carretta, T. R., & Ree, M. J. (1994). Pilot candidate selection method (PCSM): Sources of validity. *The International Journal of Aviation Psychology, 4,* 103-117.

Carretta, T. R., & Ree, M. J. (1997). *The best retest is the average: Findings and implications* (AL/HR-TP-1996-0021). Brooks Air Force Base, TX: Armstrong Laboratory Human Resources Directorate, Manpower and Personnel Research Division.

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

Ree, M. J., & Carretta, T. R. (1994). The correlation of general cognitive ability and psychomotor tracking tests. *International Journal of Selection and Assessment, 2,* 209-216.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 356-442). Washington, DC: American Council on Education.