

'AD-A094 945

RAND CORP SANTA MONICA CA
PERFORMANCE CHANGES ON MILITARY QUALIFICATION TESTS DURING THE --ETC(U)
MAR 80 W MCNAUGHT
RAND/P-6464

F/6 5/9

UNCLASSIFIED

NL

1 of 1
24
AD-A094 945



END
DATE
FILMED
3-4-81
DTIC

LEVEL

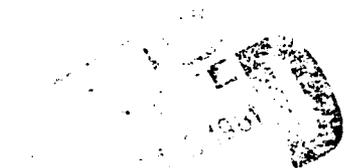
①

12) 1

PERFORMANCE CHANGES ON MILITARY QUALIFICATION TESTS DURING THE FIRST TERM OF SERVICE,

10 William McNaught

11 March 1980



14) = D/P-6464

RAND/P-6464

AD A094945

PERFORMANCE CHANGES ON MILITARY QUALIFICATION TESTS DURING THE FIRST TERM OF SERVICE

DDC FILE COPY

McNaught

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

P-6464

9 81 2 10 107

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests. Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

PERFORMANCE CHANGES ON MILITARY QUALIFICATION
TESTS DURING THE FIRST TERM OF SERVICE

There has been much recent discussion of late about the uses and abuses of military qualification tests during the enlistment process. The discussion has touched on many issues--whether military testing procedures discriminate against minorities, whether these tests provide useful indicators of success in training and on the job, and whether these tests are aptitude or achievement-based. Perhaps the most important issue is the validity of the norming of the current Armed Services Vocational Battery (ASVAB). Robert B. Pirie, Assistant Secretary for Defense for Manpower, Reserve Affairs, and Logistics has stated the ASVAB overestimates the quality of recruits in lower mental categories.*

These matters are far from academic interest. If the ASVAB test was improperly normed so that its scores are biased upwards, all our measures of the quality of manpower recruited by the All-Volunteer Force are correct. The principal argument of defenders of the All-Volunteer Force has been that its quality, usually measured by the percentage of Category IV** personnel in any entering cohort, exceeds the quality of prior military forces manned with draft procedures. Statements that "Quality, broadly defined, has not changed substantially since the removal of the draft"*** lose much of their force if this result is a mere artifact of the testing procedure.

In this paper I briefly describe the results of a large Air Force experiment conducted during 1972-1973. This experiment analyzed the

* *Army Times*, March 10, 1980.

** During the 1960s all services employed the Armed Forces Qualification Test to determine the general mental ability of all recruits. The percentile ranking of recruits is used by the services to divide all recruits into five categories as follows: I (93-100), II (62-92), III (31-61), IV (10-30), and V (0-9). Category V personnel are legally exempt from enlistment. In the early 1970s each service briefly employed their own enlistment examination. The ASVAB was introduced to standardize enlistment exams used by all services.

*** Cooper, Richard V. L., "Military Manpower and the All-Volunteer Force," The Rand Corporation, R-1450-ARPA, September 1977, p. 141.

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

the test performance of airmen after they had completed an initial term of service. Because the airmen were originally tested with the old Armed Forces Qualification Test (AFQT) and secondly with the new ASVAB, we have a unique opportunity to compare these two tests. In these results I find little evidence of bias in the ASVAB. In addition, these results strongly suggest that the AFQT scores received by recruits at enlistment are poor indicators of their later abilities to perform their military tasks. These results do not necessarily conflict with the statements of Pirie. The current difficulties have been attributed to the ASVAB as implemented in 1976, i.e., versions 5, 6, and 7. The results presented here undoubtedly relate to an earlier version of the ASVAB.

In 1972 General John Ryan, then Chief of Staff, became concerned about the pattern of Air Force reenlistment rates. Category IV personnel were then, and still are, reenlisting at much higher rates than personnel in higher categories. Some manpower planners feared this pattern would lead to declines in the quality of the career force. General Ryan ordered an experiment which retested airmen who had recently reenlisted. In his words, the experiment was to answer the question, "Is a Cat IV still a Cat IV when he reenlists?"

Twelve bases in the Air Training Command, the Strategic Air Command, and the Tactical Air Command were identified as test sites. The Military Personnel Center selected 1125 airmen at these bases who were then in their fifth and sixth years of service. All Category IV personnel at each site were then retested. Only 20 percent of Category I through III personnel were retested. Of these 1125 airmen, 1054 actually took the ASVAB test. When valid completed tests were merged with full personnel records, a useable sample of 692 airmen was obtained.

All airmen in the test had enlisted between November 1965 and November 1968. Therefore their original AFQT score was generated by the Armed Force Qualification Test itself. Consequently errors in norming the ASVAB to the original AFQT distribution should introduce spurious errors into these results.

Table 1
COMPARISON OF AFQT SCORES

Accounting Dept	✓
AFQT Score	100
ASVAB Score	100
By	<i>[Signature]</i>
Date	<i>[Date]</i>
Available for Cases	
File #	
Sheet #	
	A

Old AFQT	New AFQT					Total	Average Gain
	1	2	3	4	5		
1	5.5	1.4	---	---	---	6.9	-0.6
2	8.0	18.6	1.5	---	0.1	28.2	4.5
3	0.1	12.3	12.0	1.3	---	25.7	13.1
4	0.4	2.1	18.1	18.0	0.5	39.1	15.5
Total	14.0	34.4	31.6	19.3	0.6	100.0*	10.7

* Does not sum to 100.0 due to rounding.

Table 1 shows the distribution of the old and new scores. A substantial increase in scores occurs in every class except Category I. Category I airmen of course could not increase above their original scores since they had already scored near the maximum percentile. Just the fact that these personnel did not regress to the mean is significant. The average gains in the Category III and Category IV classes were 13.1 and 15.5 percentage points respectively.

These results, viewed uncritically, seem also to point to norming errors in the ASVAB. Nearly one-half of the airmen originally classified as mental Category III by the AFQT are classified as Category I or II on the ASVAB. A majority of original Category IV airmen increase at least one category when classified on the ASVAB. The regression analysis below offers an alternative explanation for these gains.

In these regressions I analyze only the change in scores for those originally classified in Categories III and IV. Changes in the scores of airmen originally classified as Category I or Category II are invalidated by a truncation bias. That is, we cannot measure an improvement in test performance for individuals who initially scored near the maximum value. Table 2 lists the independent variables used in the analysis. All except time in service are dummy variables. The mean number of months in service was 62.7.

Table 2
 INDEPENDENT VARIABLES USED TO ANALYZE AFQT CHANGES

Variable	Percentage in Sample
Time in Service (months)	Continuous
Race (1 if black)	18.1
Grade (1 if E5)	48.5
High School Diploma (1 if obtained since enlistment)	3.4
College Attendance (1 if college course completed since enlistment)	11.6
Career Field dummies--	
Communications/Electronics	5.1
Avionics	3.8
Aircraft Systems Maintenance	3.5
Aircraft Maintenance	16.1
Mechanical/Electrical	2.3
Structural/Pavements	5.4
Transportation	3.5
Supply	8.9
Administration	9.2
Personnel	2.3

Table 3 presents the results of my regression analysis. The changes in AFQT scores over the first enlistment term are far from random. The effects of the independent variables are surprisingly consistent over the two equations. The one very important exception is the racial dummy. Black Category IV airmen did significantly worse than their white counterparts in improving their AFQT scores over their enlistment. No such effect occurred for black Category III airmen. The variable for time in service indicates that AFQT performance improves with longevity. I will return to this point momentarily.

Table 3

REGRESSION RESULTS

	CATEGORY III	CATEGORY IV
Constant	2.192 (0.26)	0.890 (0.12)
Time in Service	.125 (0.93)	.244 (1.97)
Race	.552 (0.16)	-7.864 (-4.22)
Grade	4.007 (1.66)	6.290 (2.83)
High School Diploma	-6.594 (-0.98)	-2.631 (-0.58)
College	3.858 (0.94)	8.060 (2.50)
Communications	8.898 (0.97)	8.447 (0.80)
Avionics	3.960 (0.64)	19.476 (1.36)
Aircraft Systems	8.748 (1.29)	12.808 (2.12)
Aircraft Maintenance	5.577 (1.85)	4.189 (1.54)
Mechanical/Electrical	14.985 (1.80)	17.787 (2.70)
Structural/Pavements	1.918 (0.31)	2.882 (0.88)
Transportation	-12.4881 (-2.04)	-1.163 (-0.28)
Supply	-3.089 (-0.79)	-3.086 (-0.97)
Administration	-6.452 (-1.41)	-3.218 (-1.18)
Personnel	20.184 (1.29)	7.938 (1.21)
Degrees of Freedom	169	248
R ²	.146	.225
Standard error of estimate	15.38	14.26

Variables which reflect the motivation of airmen--those attending college and those recognized for superior performance by promotion to E5--were positively correlated with AFQT gains. The high school completion dummy is not significant in either regression. This variable's effect is negative because airmen who entered as non-high school graduates have lower scores on average than high school graduates. Thus for any given measured score, non-high school graduates are more likely to have a positive error component. This error in measurement disappears, on average, when airmen are retested.

I consider the results of the dummies for primary career field extremely important and supportive of the time in service results. Each dummy variable represents airmen with a primary Air Force Specialty Code (AFSC) in a given two digit career field. Airmen assigned in those AFSCs which require more training and populated by high quality, as judged in AFQT terms, personnel exhibit positive average AFQT increases. Conversely AFSCs which offer little training, such as Transportation, Supply, and Administration, are associated with smaller gains in AFQT scores. I was mildly surprised that the Structural/Pavements career field did not exhibit results similar to the low training career fields. Its positive coefficient is however very small and nonsignificant in both instances.

These results suggest that the observed changes in AFQT performance are the result of systematic changes, not test biases. The intercept terms of these regressions should capture any test bias. Although both intercepts are positive, their t ratios are miniscule and actual values so small as to be of little consequence.

The nature of the military enlistment tests depicted by these results is distinctly achievement, not aptitude based. Age and experience substantially increase AFQT scores. Just the time in service coefficient itself accounts for over half of the average Category III gain and all of the Category IV gain. Formal schooling, either in military training courses or in civilian colleges, and promotions are also associated with AFQT gains.

Because the ASVAB may have changed substantially between 1972 and 1976, we cannot state categorically that these results reject the

possibility of norming errors in the ASVAB as currently used in DoD enlistment screening. These results do suggest that the abilities/achievements which the AFQT measures change over time. Military planners should take care when discussing the quality of career personnel. The common usage of AFQT scores taken at the enlistment point to measure the quality of career personnel is highly dubious. It obscures much of the additional information that the DoD has gained about the performance of these individuals during their initial enlistment tour.

