AD 749 697

AD

NAVAL PERSONNEL AND TRAINING RESEARCH LABORATORY

SAN DIEGO, CALIFORNIA 92152

TECHNICAL BULLETIN STB 73-1

AUGUST 1972

AN INVESTIGATION OF POSSIBLE TEST BIAS IN THE NAVY BASIC TEST BATTERY

Patricia J. Thomas

NAVAL PERSONNEL AND TRAINING RESEARCH LABORATORY
NPTRL
SAN DIEGO CALIFORNIA

26

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) U. S. Naval Personnel and Training Research Laboratory San Diego, California 92152	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP
3. REPORT TITLE AN INVESTIGATION OF POSSIBLE TEST BIAS IN THE NAVY BASIC TEST BATTERY	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)	
5. AUTHOR(S) (First name, middle initial, last name) Patricia J. Thomas	
6. REPORT DATE August 1972	7a. TOTAL NO. OF PAGES 26 7b. NO. OF REFS 5
8a. CONTRACT OR GRANT NO. b. PROJECT NO. PF55.521.005.01.10 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) STB 73-1 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
10. DISTRIBUTION STATEMENT Approved for public release; distribution unlimited.	
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Chief of Naval Personnel (Pers-A3) Navy Department Washington, D. C. 20370

13. ABSTRACT

This research investigated whether racial bias exists in the Navy Basic Test Battery (BTB), used to assign recruits to technical schools. BTB scores and school grades were obtained for approximately 105,000 whites and 2,000 blacks attending A-Schools in 1969-1970. Sufficient numbers of blacks attended 24 schools for statistical analysis of their test scores and standardized school grades.

The findings and conclusions were as follows: (1) The means of the white and black samples were significantly different for both the school grade criterion and the predictor tests, with whites scoring higher than blacks on all variables; (2) The regression lines of each race differed significantly. If single BTB tests were used in selection, overprediction of minority performance would be somewhat more common than underprediction; (3) The tests more accurately predicted the grades of white students than of black. The selection composites were valid predictors of the performance of white students in all schools and for black students in half of the schools.

It was recommended that: (1) No general raising or lowering of test cutting scores for school selection of minority group members appears warranted; and (2) Since the tests are not as valid for blacks as for whites, it is necessary to develop improved tests and/or use different combinations of existing tests. Meanwhile, changes in selection test combinations suggested in this report should be implemented.

DD FORM 1 NOV 65 1473 (PAGE 1)

0102-014-6600

UNCLASSIFIED

Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Test bias Selection tests Minority school performance Basic Test Battery						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

UNCLASSIFIED
Security Classification

Thomas, Patricia J. An investigation of possible test bias in the Navy Basic Test Battery. San Diego: Naval Personnel and Training Research Laboratory, August 1972. (STB 73-1, PF55.521.005.01.10)

SUMMARY

Problem and Background

This investigation was undertaken to determine if there is racial bias in the Navy Basic Test Battery (BTB), which is used to assign recruits to technical school training. If the BTB were found to be biased, the extent of bias and possible means for correcting its effects were to be determined.

Approach

BTB scores and Class "A" school grades were obtained for the approximately 105,000 whites and 2,000 blacks who attended "A" Schools in 1969 and 1970. The data used were taken from the 24 schools with the largest numbers of black students. Statistical analyses were conducted of the BTB scores and standardized school grades, including a comparison of the validities, by racial group, of the selection test composites actually used in the selection of students.

Findings and Conclusions

1. The black and white samples differed significantly in their performance on both the predictor tests and on the school grade criterion. The BTB mean differences ranged from .26 to .74 standard deviation units, while the average school grade difference was .36 standard deviation, with whites scoring higher than blacks on all variables (page 4).

2. The regression lines of each of the BTB tests were significantly different for blacks and whites. In practice, combinations of tests are used for school selection. If single tests were used, neither racial group would be consistently favored by the BTB. Overprediction of minority performance would be somewhat more common than underprediction (page 6).

3. The tests were more accurate in the prediction of the grades of white students than of black students. The selection composites were significantly valid predictors of the performance of white students in all schools and for black students in half of the schools (page 14).

Recommendations

1. No raising or lowering of test cutting scores for school selection of minority group members appears warranted (page 14).

2. Since the tests are not as valid for blacks as for whites, it is necessary to develop improved tests and/or use different combinations of existing tests. Such investigations are underway. In the meantime, implementation of changes in the selection test combinations suggested in this report is recommended (page 14).

AD

AN INVESTIGATION OF POSSIBLE TEST BIAS IN THE NAVY BASIC TEST BATTERY

Patricia J. Thomas

August 1972

PF55.521.005.01.10
Technical Bulletin STB 73-1

Submitted by

B. Rimland, Ph.D., Director, Personnel Measurement Research Department

Approved by

E. I. Jones, Ph.D., Technical Director
F. L. Nelson, Captain, USN
Commanding Officer

Naval Personnel and Training Research Laboratory
San Diego, California 92152

A LABORATORY OF THE BUREAU OF NAVAL PERSONNEL

SUMMARY

Problem and Background

This investigation was undertaken to determine if there is racial bias in the Navy Basic Test Battery (BTB), which is used to assign recruits to technical school training. If the BTB were found to be biased, the extent of bias and possible means for correcting its effects were to be determined.

Approach

BTB scores and Class "A" school grades were obtained for the approximately 105,000 whites and 2,000 blacks who attended "A" Schools in 1969 and 1970. The data used were taken from the 24 schools with the largest numbers of black students. Statistical analyses were conducted of the BTB scores and standardized school grades, including a comparison of the validities, by racial group, of the selection test composites actually used in the selection of students.

Findings and Conclusions

1. The black and white samples differed significantly in their performance on both the predictor tests and on the school grade criterion. The BTB mean differences ranged from .26 to .74 standard deviation units, while the average school grade difference was .36 standard deviation, with whites scoring higher than blacks on all variables (page 4).

2. The regression lines of each of the BTB tests were significantly different for blacks and whites. In practice, combinations of tests are used for school selection. If single tests were used, neither racial group would be consistently favored by the BTB. Overprediction of minority performance would be somewhat more common than underprediction (page 6).

3. The tests were more accurate in the prediction of the grades of white students than of black students. The selection composites were significantly valid predictors of the performance of white students in all schools and for black students in half of the schools (page 14).

Recommendations

1. No raising or lowering of test cutting scores for school selection of minority group members appears warranted (page 14).

2. Since the tests are not as valid for blacks as for whites, it is necessary to develop improved tests and/or use different combinations of existing tests. Such investigations are underway. In the meantime, implementation of changes in the selection test combinations suggested in this report is recommended (page 14).

REPORT USE AND EVALUATION

Feedback from consumers is a vital element in improving products so that they better respond to specific needs. To assist the Chief of Naval Personnel in future planning, it is requested that the use and evaluation form on the reverse of this page be completed and returned. The page is preaddressed and franked; fold in thirds, seal with tape, and mail.

Department of the Navy

Postage and Fees Paid
Navy Department

Official Business

Commanding Officer
Naval Personnel and Training Research Laboratory
San Diego, California 92152

Report Title & No: AN INVESTIGATION OF POSSIBLE TEST BIAS IN THE NAVY BASIC TEST BATTERY (STB 73-1)

1. Evaluation of Report. Please check appropriate column.

FACTORS	RATING			COMMENTS
	LOW	AVE	HIGH	
Usefulness of Data				
Timeliness				
Completeness				
Technical Accuracy				
Validity of Recommendations				
Soundness of Approach				
Presentation and Style				
Other				

2. Use of Report. Please fill in answers as appropriate.

a. What are your main uses for the material contained in the report?

b. What changes would you recommend in report format to make it more useful?

c. What types of research would be most useful to you for the Chief of Naval Personnel to conduct?

d. Do you wish to remain on our distribution list?

e. Please make any general comments you feel would be helpful to us in planning our research program.

NAME: ______________________________ CODE: __________

ORGANIZATION: ______________________________

ADDRESS: ______________________________

CONTENTS

TABLES

FIGURES

AN INVESTIGATION OF POSSIBLE TEST BIAS IN THE NAVY BASIC TEST BATTERY

A. BACKGROUND AND PURPOSE

A great deal of research effort recently has been devoted to the study of possible test bias. Selection instruments used by colleges, industry, and government are being scrutinized to determine whether tests developed for use with predominantly white populations are reasonably predictive of the performance of black (or other minority) populations. In general, test bias results from inappropriately applying performance estimate equations developed on the basis of a majority sample to a minority group. Consistent underprediction of the criterion scores of minority members is referred to as negative bias. Conversely, overprediction of the performance of the minority group is referred to as positive bias.

A review of the relevant literature generally supports the conclusion that negative bias is not common. Cleary (1968) found no evidence of negative bias in her investigation of the Scholastic Aptitude Test as a predictor of grades at three colleges. O'Leary, Farr, and Bartlett (1970) conducted seven studies of predictor-criterion relationships in job situations. They concluded that test bias did exist in the majority of comparisons between blacks and whites but that the tests were as likely to favor blacks as discriminate against them. Guinn, Tupes, and Alley (1970), working with an Air Force enlisted population, investigated differences in validities for various groups. They found that the performance of blacks in technical schools was generally overpredicted; i.e., black students earned lower grades than would be expected from their test scores.

In the past, the Navy has had too few blacks in most Class "A" schools to permit investigation of whether its classification test battery, the Basic Test Battery (BTB), is discriminatory. While the absolute number of Negro enlisted men has not risen substantially over the past few years, the number of blacks assigned to schools has almost doubled.[1] This has been the result of a deliberate Navy effort to assure Class "A" school training for all black recruits who meet the minimum requirements for such training. During calendar years 1969-1970, the period with which this report is concerned, blacks were

[1]Black representation in Class "A" schools is increasing rapidly. Since 1 February 1972 classifiers have been directed to assign all black recruits who are school eligible to school training. The previous policy was, in essence, to select men for school from the pool of eligibles on the basis of test score, minimization of travel costs, and similar factors, without regard to color.

sufficiently represented among the graduates of 24 Class "A" schools for inclusion in a bi-racial validity study of the BTB.

B. PROCEDURE

1. Test Bias

The problem of determining if a test is biased is complicated by the number of ways in which a test may be discriminatory. This study will concentrate on two commonly accepted definitions of bias, or lack of bias. The first is that of Cleary (1968), who stated, "A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup." Statistically, this type of bias is investigated by testing the slopes and intercepts of the regression lines for the majority and minority populations to determine whether they differ significantly. The method used for performing these tests was developed by Gulliksen and Wilks (1950). The second definition of discrimination investigated, involving test fairness, is that of the Department of Labor whose regulations must be complied with by all federal contractors. In Title 41 of the Code of Federal Regulations (1971) the following directions for assessing the validity of a selection test are given: "The relationship should be sufficiently high as to have a probability of no more than 1 to 20 to have occurred by chance . . . . A test which is differentially valid may be used in groups for which it is valid but not for those in which it is not valid." To determine whether the recruit classification tests would comply with this standard, the BTB selection composites were validated against final grades in Navy schools separately for black and white samples.

2. Sample

Data routinely gathered for graduates and disenrollees from Class "A" schools formed the basis of the sample. BTB scores and racial information were obtained for students completing school training in 1969 and 1970. The data were sorted by race and school code to determine which schools had sufficiently large numbers of blacks for a bi-racial analysis of possible selection test bias. Twenty-four schools (out of approximately 140) which had at least 19 black students among their graduates or academic disenrollees were selected. The total number of white students, combined across all "A" schools, with complete predictor and criterion variables was 104,683. The blacks numbered 2067.[2] The records

---

[2] Although the representation of blacks in Class "A" schools is very low (2%), it does not appear that they were being discriminated against by the recruit classification process, since the blacks who were assigned to schools scored significantly lower on all BTB tests than did whites (see Table 1).

of blacks had not been isolated in previous BTB studies because of their small representation in the school samples and because the problem of possible test bias was not a salient issue. Now, however, since Title 41 has shifted the burden of proof of nondiscrimination to the employer, the military services have undertaken to determine whether or not their selection tests are biased.

3. Variables

a. Basic Test Battery (BTB). Six of the basic and special tests in the Navy battery were used as predictors. Scores are reported as Navy Standard Scores having a mean of about 50 and a standard deviation of about 10 for an unrestricted recruit population. The tests are:

(1) General Classification Test (GCT)--consisting of 60 verbal analogy and 40 sentence completion items with a single 35-minute time limit.

(2) Arithmetic Reasoning Test (ARI)--consisting of 30 arithmetic reasoning items with a 35-minute time limit.

(3) Mechanical Test (MECH)--consisting of two separately timed 50-item subtests yielding a single score. The tool knowledge section has a 10-minute time limit and the mechanical comprehension section has a 25-minute time limit.

(4) Clerical Test (CLER)--consisting of 100 number matching items. This highly speeded test has a 5-minute time limit.

(5) Shop Practices Test (SP)--consisting of 30 items with a 17-minute time limit.

(6) Electronics Technician Selection Test (ETST)--consisting of three separately timed sections: Mathematics (20 items in 25 minutes); Science (20 items in 15 minutes); and Electricity and Radio (30 items in 20 minutes).

b. Armed Forces Qualification Test (AFQT). The AFQT, administered to all Selective Service registrants, is used as a measure of general ability. Scores are reported as percentiles. A minimum percentile score of 10 was established by the Congress to indicate mental fitness for military training. The aptitude areas covered by the 100 items in the AFQT are verbal, arithmetic reasoning, tool functions, and spatial relations. A 50-minute time limit is used.

c. Final School Grade (FSG). The grade given by the Class "A" schools upon graduation or disenrollment was used as the criterion. It is most commonly a weighted sum of grades earned on daily and/or weekly quizzes, measures of practical proficiency, and the score on the final examination. FSG ranges from about 35 to 99 in its raw form. It was standardized to a mean of 50 and a standard deviation of 10 within each school for some of the analyses in this study.

4. Analysis of Data

Means, standard deviations, and correlations among the test variables and standardized FSG were computed for the two racial samples combined across all samples. The significance of the differences between paired statistics for the black and white groups was determined. The regression lines for each BTB test and AFQT were plotted separately by race and tested for differences in errors of estimate, slopes, and intercepts, using the method of Gulliksen and Wilks (1950).

In practice, qualification for a Navy school is not determined by a score on a single BTB test. Instead, a summed combination of two or three BTB tests is used in the classification decision. Thus, the most relevant statistic for judging the effectiveness of the battery in school selection is the correlation between this composite and FSG. These correlations were computed separately for each race and tested for significance as required by Title 41. The differences between the validities for blacks and whites within each rating were also tested.

C. RESULTS AND DISCUSSION

1. Differences Between Racial Means, Standard Deviations and Validities

Table 1 presents BTB, AFQT, and FSG statistics for the total white and black samples, both of which were combined across all schools submitting data in order to maximize the size of the black sample. All of the mean scores differed significantly, with the whites consistently performing higher both on the tests and in schools. With one exception, that of CLER, the tests were also significantly more valid for the whites, even though the standard deviations of the variables were very similar for the two races. Although these results clearly show that the BTB and the AFQT are better predictors of the school grades of white enlisted men than of black, the test validities for the blacks were significantly different from zero.

The interpretation of test statistics combined across all schools has certain limitations. Navy students are selected using some, but not all, of the variables under consideration. Therefore, the amount of restriction in the range of each variable is different and, of course, has a differing effect on the magnitude of the validity coefficients. In addition, using all of the test scores of each individual usually has a depressing effect upon the validities of the tests <u>not</u> used in his selection. For example, MECH may have a low correlation with grades in Postal Clerk school, a substantial correlation with grades in Aviation Ordnanceman school, and an intermediate validity across all schools.

TABLE 1

Means, Standard Deviations and Validities of the BTB and AFQT for Black and White Samples Combined Across All "A" Schools

Variable	White (N=104,683)			Black (N=2067)			Difference Between Validities	Difference Between Means
	Mean	Standard Deviation	Validity	Mean	Standard Deviation	Validity		
GCT	59.61	7.11	.39**	54.54	6.89	.25**	.14**	5.07**
ARI	57.70	6.99	.37**	52.37	6.74	.26**	.11**	5.33**
MECH	53.94	7.41	.21**	46.57	7.15	.15**	.06**	7.37**
CLER	53.05	8.61	.13**	50.49	8.79	.10**	.03	2.56**
SHOP	55.57	6.99	.24**	48.85	7.32	.15**	.09**	6.72**
ETST	60.26	8.04	.41**	55.30	8.72	.28**	.13**	4.96**
AFQT	75.00	16.24	.34**	57.34	18.36	.26**	.08**	17.66**
Standardized Final School Grade	50.10	9.81	--	46.51	9.44	--	--	3.59**

$*p < .05$

$**p < .01$

In spite of this limitation, the BTB means, standard deviations, and validities were computed for the total sample in each racial group in order to obtain a large enough black sample for stable statistics.

2. Differences Between Regression Lines

Figures 1 through 7 show the regression lines for the majority and minority populations taken separately; i.e. the relationships between the final school grade and test score for each race. They provide information concerning the bias that may be occurring, if any. The regression lines for MECH and CLER show consistent positive bias; that is, the school grades of blacks would be predicted to be somewhat higher when based on a majority sample than when based on a minority sample, for all score levels. The remaining figures show the grades of blacks scoring low on the tests to be underpredicted by white regression equations and the opposite for blacks with high test scores. For the most part the two regression lines cross below the mean test score (indicated by the dot on the regression line) of the minority sample. Thus, in the case of blacks, overprediction (predicted performance being higher than actual performance) is more common than underprediction. On the AFQT, however, the lines cross just above the mean test score of the blacks (57th percentile) so that over- and underprediction occur with almost equal frequency.

The Gulliksen and Wilks (1950) chi-square tests of the significance of the differences between the regression lines are reported on the figures. This method tests three hypotheses concerning whether the populations from which the two groups were drawn can reasonably be said to be different. Hypothesis (1) is that the standard errors of estimate (population variances) are equal. Assuming hypothesis (1) to be true, hypothesis (2) is that the regression lines are parallel. Finally, assuming that hypotheses (1) and (2) are true, hypothesis (3) is that the regression lines are identical (or have equal intercepts). These hypotheses are tested sequentially and, when one is rejected, no further tests need be made to show that the two populations differ signficantly.

Hypothesis (1) was rejected for ARI alone, since the remaining six tests showed no significant difference between the errors of estimate of the black and white populations. Hypothesis (2) was rejected for all tests except CLER, which subsequently showed significant differences between the intercepts of the two races. The results obtained on all seven aptitude tests thus demonstrated that the two racial populations were dissimilar with respect to the relationships between the tests and the criterion variable. As a matter of fact, the regression lines in each case are significantly different for whites and blacks.

3. Comparison Between Selector Score Validities

The most meaningful type of bias analysis is one which studies the tests as they are actually used in selection. This involves looking at the validities of the test combinations, as predictors of performance

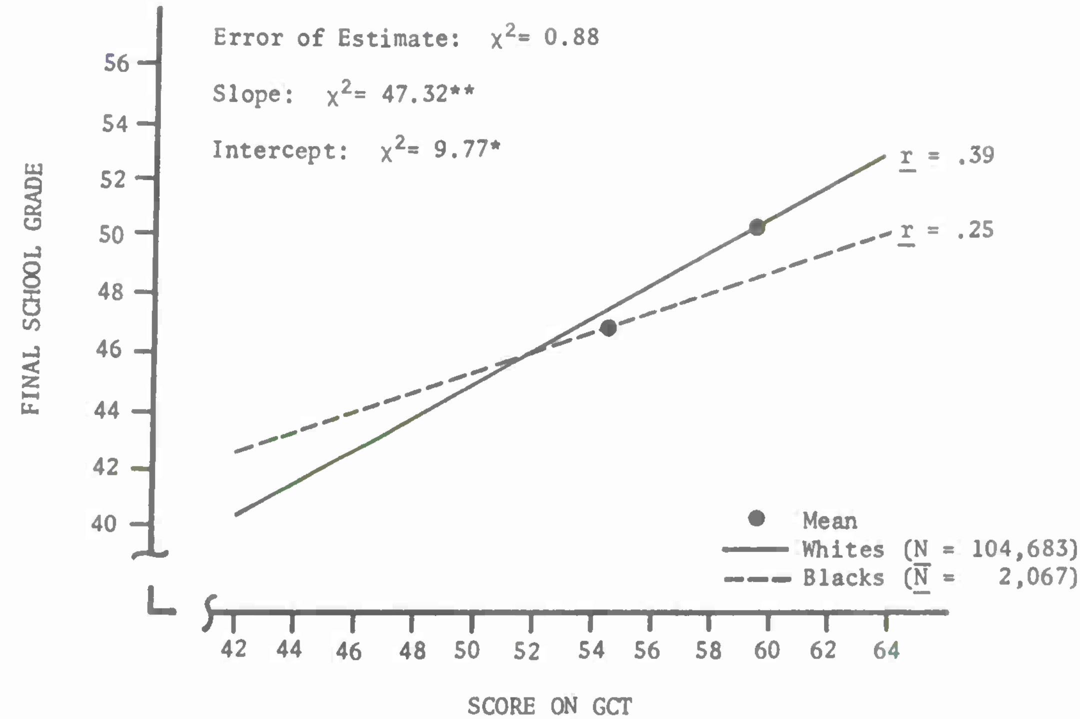

Fig. 1. Regression lines of white and black samples for the General Classification Test.

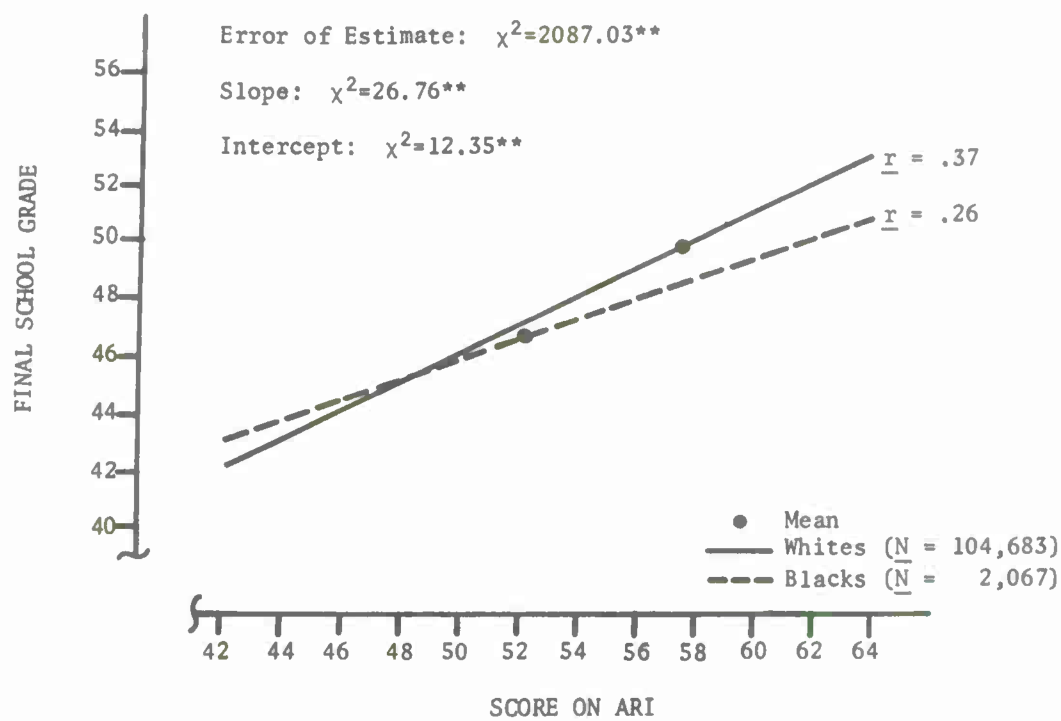

Fig. 2. Regression lines of white and black samples for the Arithmetic Reasoning Test.

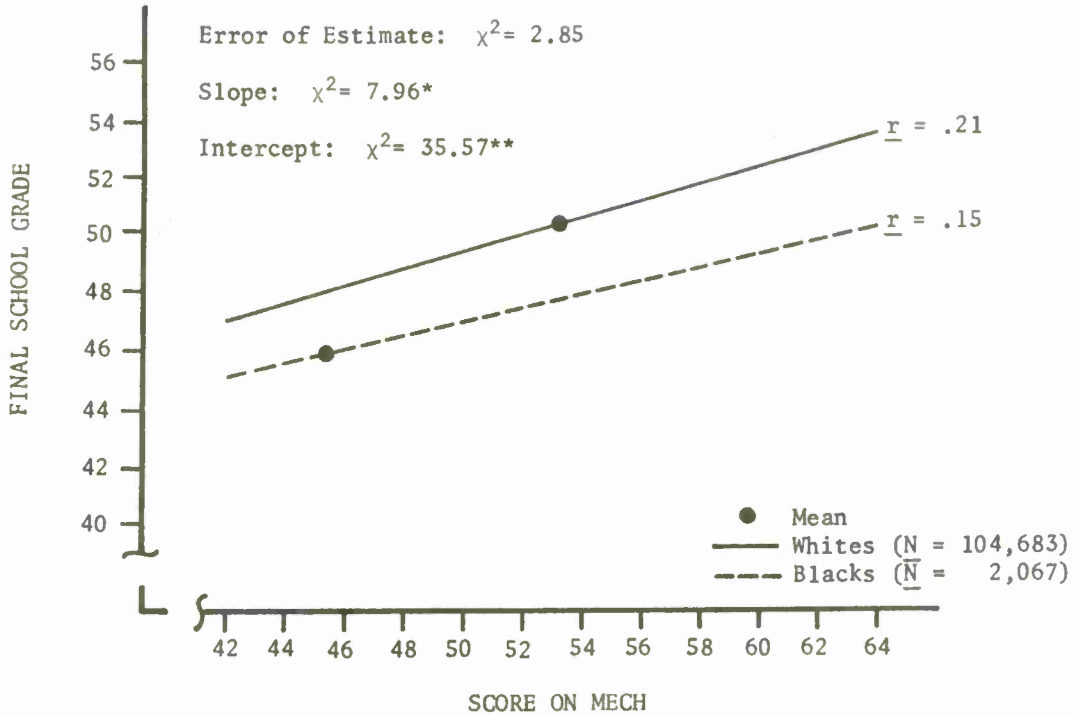

Fig. 3. Regression lines of white and black samples for the Mechanical Test.

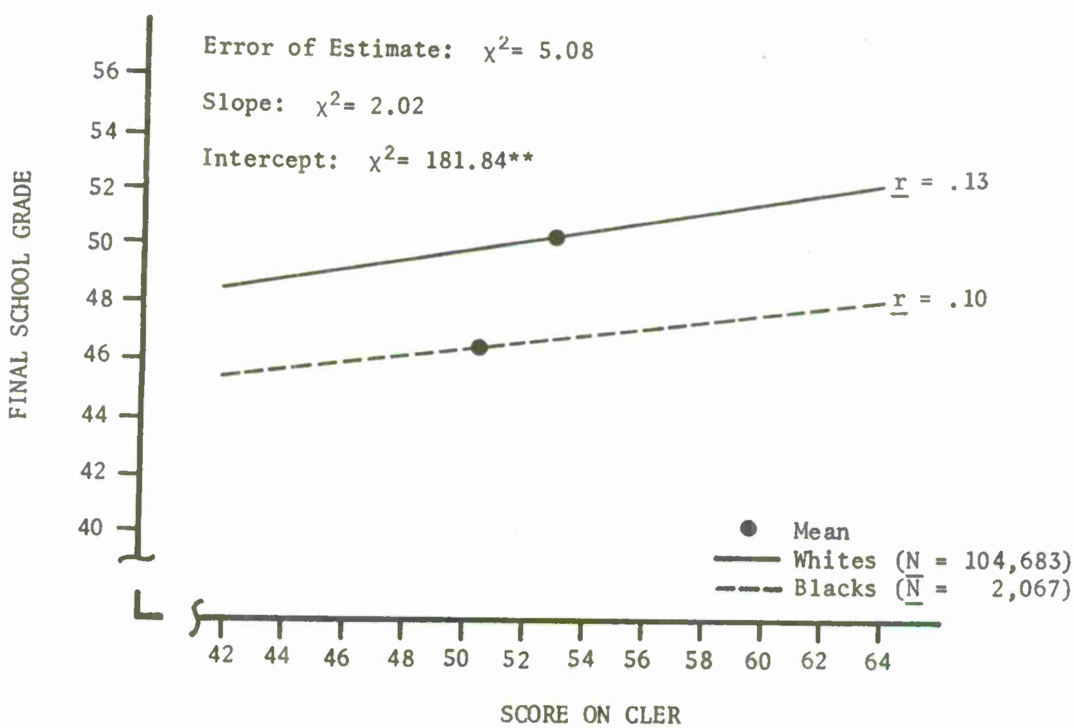

Fig. 4. Regression lines of white and black samples for the Clerical Test.

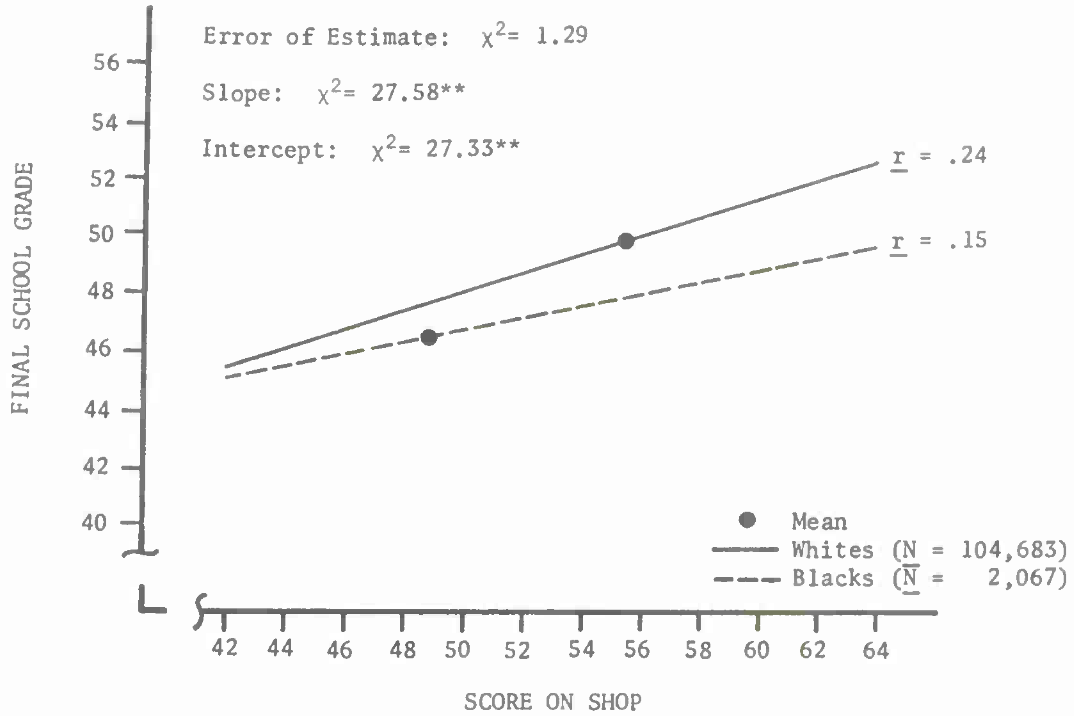

Fig. 5. Regression lines of white and black samples for the Shop Practice Test.

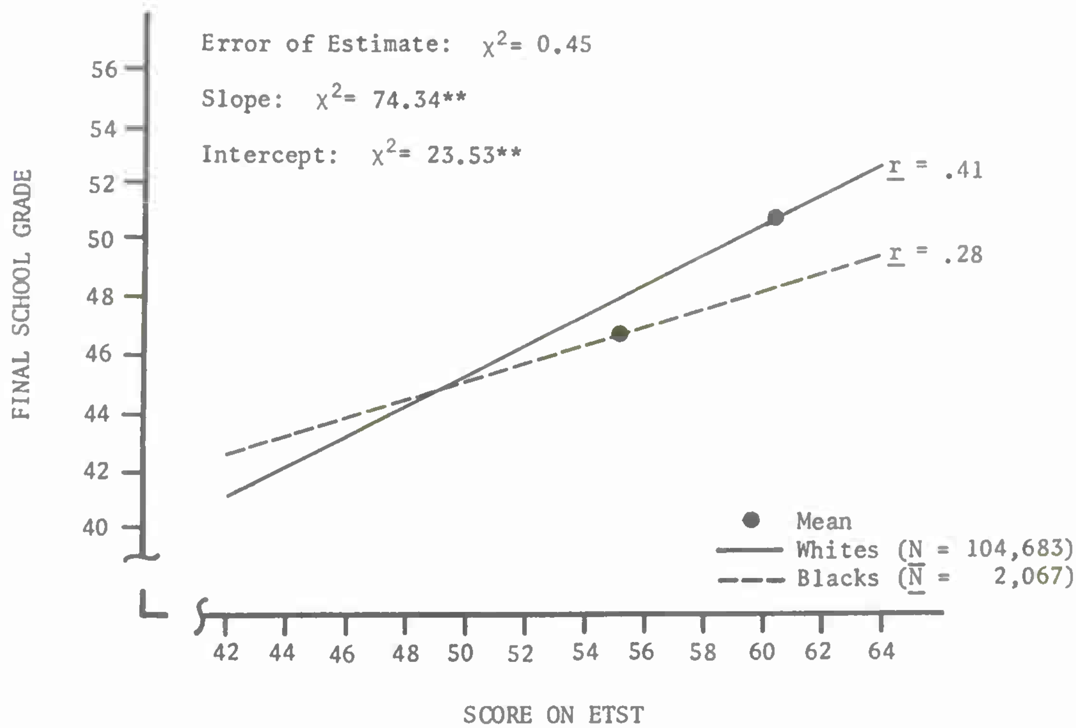

Fig. 6. Regression lines of white and black samples for the Electronics Technician Selection Test.

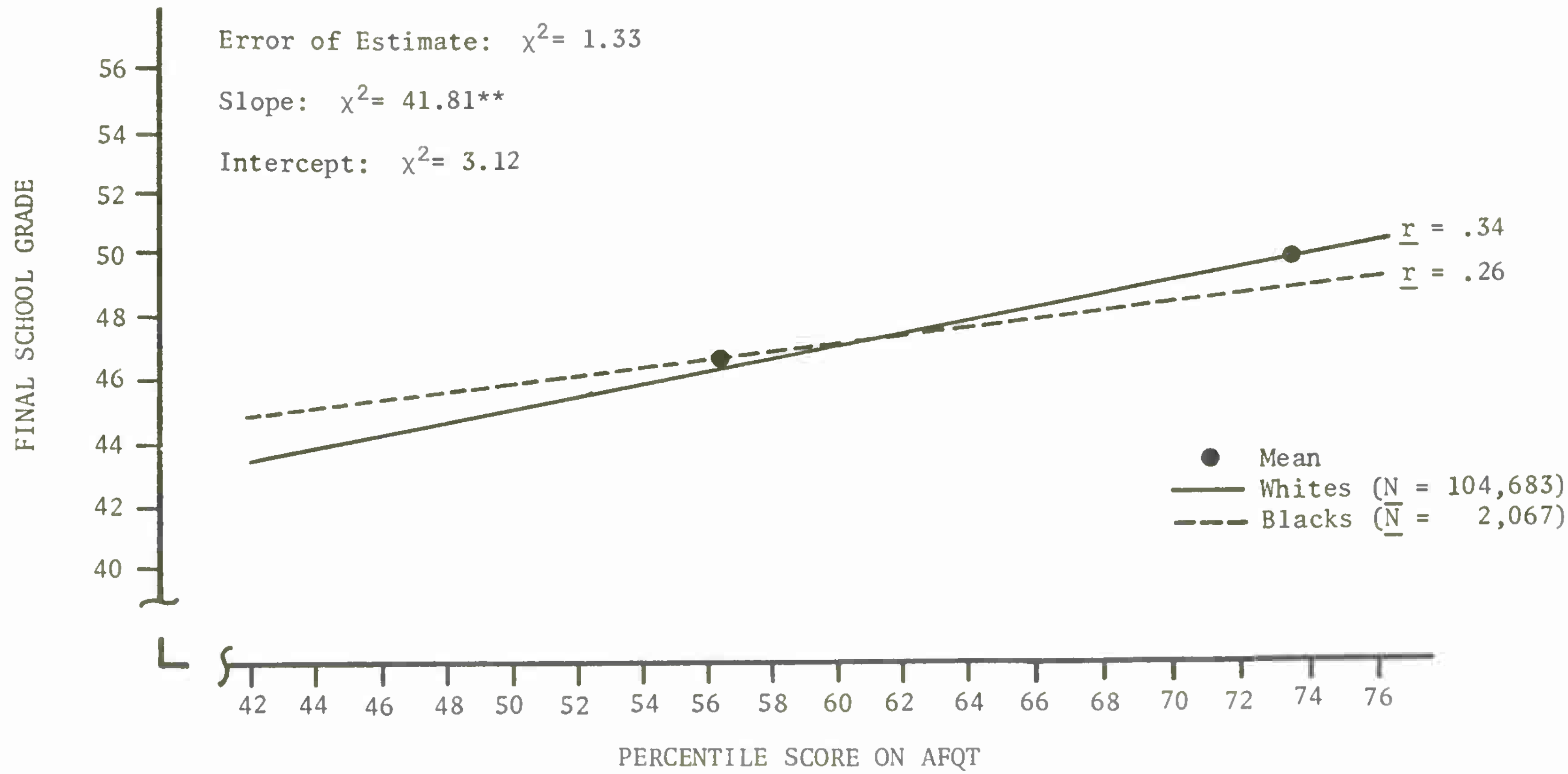

Fig. 7. Regression lines of white and black samples for the Armed Forces Qualification Test.

in the relevant schools, for black and white samples separately. Only 22 of the 24 Class "A" schools were used in this analysis because the two Basic Electricity & Electronics schools had varying selectors. Schools for the same rating were combined since the end-product of the classification system is assignment to training in a rating, not to a specific school.

Table 2 presents a comparison of the uncorrected correlations between school selectors and school grades for black and white students (corrected correlations are presented in Table 4 in the Appendix). These are linear-summed validities, rather than multiple correlations, because in practice test scores are simply added together to determine school eligibility (with the exception of ARI+2ETST selector in which a weight of two is applied to one test). The operational selector composites were predictive of the school performance of the white students at the .01 level of significance for every rating in the analysis. These same selectors failed to predict the grades of black students above chance levels in nine of the 18 ratings.[3] From this analysis it appears that these test combinations do not meet the requirements of Title 41 for the minority group. However, it might be feasible to use other BTB test composites for school selection. The most valid two composites for each school in which the operational selector failed to yield significant correlations with the criterion are presented in Table 3. In all nine ratings, prediction could be improved by using these suggested combinations, significantly so for six ratings. In four cases the suggested combinations also raise the validities for whites appreciably, indicating that the adoption of these composites would be beneficial. In the other five cases, the possibility of using the alternate composites only for blacks needs to be considered. This would require constructing conversion tables and making other adaptations to the classification system.

The criterion used in this report was final grade earned in Class "A" school. While training grades, as an intermediate criterion, are recognized as being less crucial than performance on the job, they nevertheless constitute a relevant criterion; for in the Navy, as in civilian life, successful completion of training is a job entry requirement.

It is sometimes argued that level of academic achievement usually shows only a marginal relationship to level of job performance. This

---

[3]As in statistical testing in general, the failure to reject the null hypothesis of no correlation does not establish the fact of zero correlation. The failure to find a significant relationship between the selection composites and FSG is partly a function both of the smaller number of black students and the low correlations. If the significance levels were recomputed using the same correlations for blacks but assuming the _Ns_ to equal those of whites at the same schools, the operational selectors would fail to predict above chance levels in only four of the 18 ratings.

TABLE 2

Validities of Linear Sum Composites
for White and Black Samples

Rating	Selector	White N	White r	Black N	Black r	Difference
ADR	GCT+MECH+SP	3009	.41**	50	.30*	.11
AE	GCT+MECH+SP	4063	.47**	75	.25*	.22*
AM (2)	GCT+MECH+SP	3297	.45**	46	.34*	.11
AO	GCT+MECH+SP	2613	.42**	45	.13	.29*
AV	ARI+2ETST	8319	.61**	128	.35**	.26**
AVI	GCT+MECH+SP	7003	.53**	122	.48**	.05
AZ	GCT+ARI	554	.43**	19	.56*	-.13
CTR	GCT+ARI	1775	.37**	24	.01	.36
CYN	GCT+CLER	924	.47**	54	.25	.32
DT	GCT+ARI	996	.55**	54	.19	.36**
ET (2)	ARI+2ETST	6162	.60**	76	.55**	.05
HM (2)	GCT+ARI	10,970	.63**	571	.37**	.26**
PC	GCT+CLER	176	.32**	20	.58**	-.26
PE	ARI+2ETST	2004	.55**	42	-.02	.57**
QM	GCT+CLER	850	.41**	30	.00	.41*
RM	GCT+ARI	4921	.38**	87	.17	.21*
SK (2)	GCT+ARI	2351	.50**	57	.21	.29*
SM	GCT+CLER	912	.29**	39	-.01	.30
Means			.53**		.33**	.20**

Note.--(2) indicates that the data for two "A" Schools were combined.

$*p < .05$

$**p < .01$

TABLE 3

Alternative Selectors for Schools in Which Black Validities Were Not Significant

Rating	Current Selector	Black *r*	White *r*	Best Two-Test Composite[a]	Black *r*	White *r*	Next Best Two-Test Composite[a]	Black *r*	White *r*
AO	GCT+MECH+SP	.13	.42**	ARI+ETST	.32*	.49**	CLER+ETST	.27	.40**
CTR	GCT+ARI	.01	.37**	GCT+ETST	.26	.34**	GCT+CLER	.16	.29**
CYN[1]	GCT+CLER	.25	.47**	ARI+ETST	.50**	.56**	GCT+ARI	.47**	.58**
DT	GCT+ARI	.19	.55**	GCT+CLER	.40**	.46**	CLER+ETST	.31*	.47**
PE	ARI+2ETST	-.02	.55**	MECH+SP	.13	.25**	MECH+CLER	.11	.22**
QM	GCT+CLER[2]	.00	.41**	ARI+SP	.42*	.51**	GCT+SP	.34	.43**
RM	GCT+ARI	.17	.38**	CLER+ETST	.30**	.32**	ARI+ETST	.26*	.36**
SK	GCT+ARI	.21	.50**	GCT+CLER	.40**	.43**	CLER+ETST	.38**	.43**
SM	GCT+CLER	-.01	.29**	GCT+ARI	.25	.40**	GCT+ETST	.13	.34**

Note.--

[1]CYN rating was disestablished after present data were collected.

[2]The operational selector for QM school has recently been changed to GCT+ARI (*r* = .30 for blacks; .62** for whites).

[a]Based on black samples only.

*$p < .05$

**$p < .01$

relative lack of relationship is largely attributable to the limitations inherent in rating scales, the traditional way of measuring performance on the job. Rating scales, as well as other available methods of measuring on-job performance, suffer from many serious deficiencies, including low reliability, subjectivity, incomplete coverage of job duties, and lack of standardization between billets.

This report is concerned with the validity of the Navy's Basic Test Battery, which measures aptitude for school training. The sample consisted of students from a variety of schools being trained for disparate ratings. Final school grades were deemed the appropriate criterion for the BTB and a more reliable criterion than the operational performance rating. While it is recognized that the ultimate goal of the selection process is choosing men who can adequately perform the job, completion of school training is a hurdle that must be cleared.

D. CONCLUSIONS

The black and white samples differed significantly in their performance on both the predictor tests and on the school grade criterion. The BTB mean differences ranged from .26 to .74 standard deviation units, while the average school grade difference was .36 standard deviation, with whites scoring higher than blacks on all variables.

Six of the seven test validities were significantly different and the hypothesis that the two samples were drawn from a homogeneous population was rejected. In the strict statistical sense adopted by Cleary, it has been demonstrated that significant nonzero errors of prediction would be made for the minority population. However, the result of these errors would be inconsistent and overprediction of minority performance would be a more common occurrence than underprediction. Thus, no lowering or raising of cutting scores for minority members appears warranted.

On the practical and legal question of the validity of the school selection composites, it was shown that for half of the ratings the selectors failed to predict the performance of black students at the .05 level of significance. However, if other combinations of BTB tests were used as selectors to these schools, the predictive validity could be raised appreciably in two-thirds of the cases.

The possible existence of selection bias, which has not been ruled out by these findings, makes further investigation imperative. Under instructions from the Chief of Naval Operations, many more blacks are being assigned to formal school training. If analysis of larger samples confirms the apparent differences between validities for blacks and whites, improved selection composites will have to be employed for minority recruits. In the meantime, it is recommended that alternate selection composites be implemented for school training in the Aviation

Ordnanceman, Quartermaster, and Signalman ratings. These revised composites, identified in Table 3, were found to improve the prediction of school performance for members of both the majority and minority races.

The Communications Yeoman rating, which Table 3 also indicates as potentially benefiting from a change in selection test composites, is not included in the foregoing recommendation because it has recently been disestablished. Although the recommendation to use ARI+ETST in selection for the AO rating appears to have little logical basis, it is a more valid composite than the current selector used in recruit classification, particularly for blacks. This finding seems stable, since BTB validation studies of data collected during 1964-1966 and 1966-1968 also showed ARI+ETST to be equally or more valid than GCT+MECH+SP in predicting grades in Aviation Ordnanceman Class "A" school. An analysis of the ETST is planned to determine why it predicts well in nonelectronic ratings. Perhaps certain subtests or item types can be extracted for more extensive use. Since men who score high on the ETST are in great demand for electronic training, selection composites for nonelectronic schools typically do not include ETST.

REFERENCES

Cleary, T. Anne. Test bias: prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5 (2), 115-124.

Department of Labor. Proposed employment opportunity program. Federal Register, 1971, 36, No. 77.

Guinn, Nancy, Tupes, E. C. & Alley, W. E. Cultural subgroup differences in relationships between Air Force aptitude composites and training criteria. AFHRL-TR-70-35. Lackland AFB, Texas: Personnel Research Division, Air Force Human Resources Laboratory, September 1970.

Gulliksen, H. & Wilks, S. S. Regression tests for several samples. Psychometrika, 1950, 15 (2), 91-114.

O'Leary, B. S., Farr, J. L. & Bartlett, C. J. Ethnic group membership as a moderator of job performance. Office of Naval Research, Technical Report No. 1, April 1970. (American Institutes for Research, Contract No. N00014-68-C-0341)

APPENDIX

CORRECTIONS FOR RESTRICTION OF RANGE

Since school assignment is usually contingent upon achieving a minimum score on a combination of tests, restriction in the range of test scores is present in school samples. This is evident in that the mean selector scores of students are higher than those of the general recruit population and the standard deviations of these tests are generally smaller for student samples. Such restriction may be expected to result in lower test validities for school samples than for more heterogeneous recruit samples. Therefore, validities derived from school samples are usually statistically corrected to yield estimates of test validities for the full-range recruit population as well as to express validities on a common base so that they can be compared across schools and for different time periods.

Several questions arise concerning the means of correcting for restriction of range of two different racial samples. If the samples were drawn from statistically different populations, which full-range population values should be used in making corrections? If the decision is to use two correction populations, one relevant to each of the heterogeneous samples, can the resultant corrected correlations be considered statistically comparable? This procedure loses the value of correcting to a common base. It is quite possible that one sample may be a more restricted subsample of its population than is the other sample. Therefore, the corrected correlations of the former would show a greater increment over the uncorrected correlations than would those of the latter.

Throughout this analysis, the practice of reporting and interpreting uncorrected correlations has been adopted. However, the reader may subscribe to the position that since selection occurs within a total recruit population containing a minority of blacks, a matrix based on an unrestricted sample of recruits is the relevant population for restriction of range corrections, in spite of the possible error involved. Therefore, Table 4, concerning the crucial question of bias in operationa selection composites, was prepared to permit comparison with Table 2. The mean increase in selector validities with white samples was .11 correlation points; with black samples, it was .08 correlation points.

APPENDIX (continued)

TABLE 4

Difference Between Corrected Validities of Selector Scores for White and Black Samples

Rating	Selector	White N	White r_c	Black N	Black r_c	Difference
ADR	GCT+MECH+SP	3009	.49	50	.35	.14
AE	GCT+MECH+SP	4063	.55	75	.29	.26
AM	GCT+MECH+SP	3297	.51	46	.42	.09
AO	GCT+MECH+SP	2613	.49	45	.16	.33
AV	ARI+2ETST	8319	.82	128	.54	.28
AVI	GCT+MECH+SP	7003	.60	122	.53	.07
AZ	GCT+ARI	554	.51	19	.57	-.06
CTR	GCT+ARI	1775	.44	24	.02	.42
CYN	GCT+CLER	924	.48	54	.26	.22
DT	GCT+ARI	996	.59	54	.24	.35
ET	ARI+2ETST	6162	.82	76	.76	.06
HM	GCT+ARI	10,970	.67	571	.46	.21
PC	GCT+CLER	176	.36	20	.53	-.17
PE	ARI+2ETST	2004	.74	42	-.03	.77
QM	GCT+CLER	850	.48	30	.04	.44
RM	GCT+ARI	4921	.44	87	.21	.23
SK	GCT+ARI	2351	.57	57	.25	.32
SM	GCT+CLER	912	.32	39	.09	.23
Means			.64		.44	.20

DISTRIBUTION LIST

Chief of Naval Operations (OP-039) (OP-059) (OP-098T) (OP-98TL) (OP-099) (OP-964) (OP-987F)
Chief of Naval Personnel (Pers-1) (Pers-11b w/4 Library Index Cards) (Pers-A) (Pers-A1) (Pers-A2) (Pers-A3 w/4 Library Index Cards) (Pers-A23) (Pers-A31) (Pers-A32) (Pers-A4) (Pers-B) (Pers-B1) (Pers-B2t) (Pers-B2211) (Pers-B2232) (Pers-P) (Pers-P1) (Pers-P11)
Office of Naval Research (Code 458) (2)
Assistant Secretary of Navy for Manpower and Reserve Affairs (2)
Office of Assistant Secretary of Defense (M&RA)
Commander, Navy Recruiting Command (Codes 00; 015; 20; 312; 314)
Chief, Naval Technical Training (Code 75), NAS Memphis
Chief of Naval Training (Codes 017; N5), NAS Pensacola
Commanding Officer, Personnel Research and Development Laboratory (3)
Commanding Officer and Technical Director, Naval Electronics Laboratory Center, San Diego (2)
Commanding Officer, Navy Medical Neuropsychiatric Research Unit, San Diego
Commanding Officer, Naval Aerospace Medical Institute, Pensacola
Commanding Officer, Naval Examining Center, Great Lakes
Commanding Officer, Naval Submarine Medical Center, Groton
Commanding Officer, Naval Medical Research Institute NMMC, Bethesda
Commander, Personnel Research Division, Air Force Human Resources Laboratory, Lackland AFB, Texas
Commandant (Code B-5), U. S. Coast Guard, Washington, D. C.
Assistant Chief of Staff (G-1), U. S. Marine Corps
U. S. Army Enlisted Evaluation Center Library, Fort Benjamin Harrison (2)
Headquarters, U. S. Marine Corps (Code A01B), Washington, D. C.
U. S. Army Behavior and Systems Research Laboratory, Arlington, VA
Superintendent, Naval Postgraduate School, Monterey
U. S. Military Academy, West Point
Superintendent, Naval Academy, Annapolis
Superintendent, U. S. Air Force Academy, Colorado
National Research Council, Division of Anthropology and Psychology, Washington, D. C.
National Science Foundation, Washington, D. C.
Director, Defense Documentation Center, Alexandria, VA (12)
Center for Naval Analyses, Arlington, VA