NPRDC TR 80-35

September 1980

VALIDITY OF A BATTERY OF EXPERIMENTAL TESTS IN PREDICTING PERFORMANCE OF NAVY PROJECT 100,000 PERSONNEL

Charles H. Cory
Nancy E. Neffson
Bernard Rimland

Reviewed by
Martin F. Wiskoff

Released by
James F. Kelly, Jr.
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(14)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPRDC-TR-80-35	2. GOVT ACCESSION NO.: AD-A091243
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): (6) VALIDITY OF A BATTERY OF EXPERIMENTAL TESTS IN PREDICTING PERFORMANCE OF NAVY PROJECT 100,000 PERSONNEL.	5. TYPE OF REPORT & PERIOD COVERED: (9) Final Report.
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): (15) Charles H. Cory, Nancy E. Neffson, Bernard Rimland	8. CONTRACT OR GRANT NUMBER(s): (16) F55521
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Navy Personnel Research and Development Center, San Diego, California 92152	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (17) 62763N, ZF55 521 030 01.01
11. CONTROLLING OFFICE NAME AND ADDRESS: (11) Navy Personnel Research and Development Center, San Diego, California 92152	12. REPORT DATE: Sept[illegible] 1980
	13. NUMBER OF PAGES: 64
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 67	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Project 100,000	Personnel Retention
Mental Group IVs	Culture Fair Tests
Personnel Selection	Personnel Promotion
Job Performance	Test Bias

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes results of a four-phase study that originated as part of the Project 100,000 research effort. The purpose of the study was to develop "culture fair" aptitude tests that would permit the Navy to identify potentially successful recruits from those who scored low on conventional tests.

Nineteen experimental tests/questionnaires were developed to measure practical (as opposed to academic) mental abilities. The experimental —>

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

3477/12

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

instruments were divided into four batteries, each of which was administered to a separate sample ranging in size from 5,000 to 12,000 recruits. The instruments were validated against supervisory performance ratings, rating progression, and retention criteria for sample members. Separate analyses were done for Mental Level IVs, Blacks, and for apprenticeship level (non-rated, undesignated strikers) and technical rating groups.

The findings were generally negative. With only a few exceptions, the experimental tests were not valid predictors of on-job performance for any of the subgroups studied, and were less valid than the conventional tests for predicting either job performance or rating advancement. Also, because of the wide variety of "culture fair" tests evaluated, it is unlikely that paper-and-pencil tests can be found that will identify previously overlooked aptitudes in low-ability populations. A number of by-product findings of potential value to those interested in optimizing the utilization of low aptitude personnel were provided from the analyses.

The findings were communicated to Navy and DoD officials when Project 100,000 was terminated; the present report provides a record of these efforts to facilitate future research in the problem area.

FOREWORD

This research and development was performed in response to a Department of Defense request initiated in 1967 to develop and evaluate a series of experimental tests that would be more appropriate than operational battery tests for selecting low mental ability personnel. The findings were communicated to the sponsors upon the completion of the work. The general tenor of the results was also summarized in AFHRL TR 76-69, an annotated bibliography pertaining to low aptitude personnel. A formal report is being published at this time because of current interest in lower ability personnel in the military.

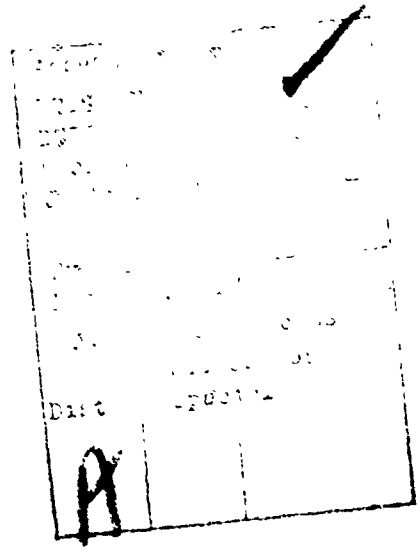
This report summarizes an extensive effort that was carried out over a 7-year time span, commencing in 1967, with important contributions from many people. Dr. Bernard Rimland and Mr. Edmund Thomas were responsible for planning the study, developing a number of the experimental tests, conducting the experimental testing, and collecting the initial criterion data. Ms. Nancy Neffson and Mr. James Stapleton carried out most of the extensive data processing efforts. Dr. Charles Cory collected the bulk of the criterion data, supervised the data analysis, and wrote the initial draft of the report.

Appreciation is expressed to the Recruit Training Command, San Diego for providing the subjects and the personnel to administer the batteries of experimental tests over a 3-year time span. Additionally, the assistance received from individual commands and supervisors throughout the fleet in completing and returning on-job performance follow-up questionnaires on the personnel in the study is deeply appreciated.

The results are intended for use primarily by cognizant officers in the Naval Military Personnel Command and in the Navy Recruiting Command.

JAMES F. KELLY, JR.
Commanding Officer

JAMES J. REGAN
Technical Director



SUMMARY

Problem

Operational tests used for Navy selection were developed to predict scholastic performance in Class "A" School rather than to measure practical abilities. Thus, use of these tests may result in exclusion of low mental ability personnel who could be used effectively and in greater numbers by the Navy. Also, it was thought that these tests might not be effective in predicting performance for minorities, particularly Blacks.

Objective

The objective of this effort was to develop and validate a comprehensive series of experimental tests that would improve the Navy's ability to discover and utilize heretofore untapped talent in its enlisted input, in regard to both low mental level and minority personnel.

Approach

Nineteen experimental tests/questionnaires were developed. The experimental instruments were divided into four batteries, each of which was then administered to samples ranging in size from 5,000 to 12,000 recruits. The men tested were followed up 21 months later by collecting supervisory performance ratings, rating progression, and retention criteria. Separate analyses were done for Mental Level IVs, for Blacks, and for apprenticeship level (nonrated, undesignated strikers) and technical rating groups.

Findings

In general, the findings were negative. With only a few exceptions, the experimental tests did not prove to be valid predictors of on-job performance for any of the subgroups studied. Further, they were less valid than the conventional tests for predicting either job performance or rating advancement.

Conclusions

1. Because of the wide variety of "culture fair" tests evaluated, it seems unlikely that paper-and-pencil tests can be found that will identify previously overlooked aptitudes in low-ability populations.

2. Analyses of the data collected in this research provided a number of by-product findings of potential interest to those involved in optimizing the utilization of low-aptitude personnel. These are summarized on pages 41-45.

Recommendations

1. None of the experimental tests included in the present research should be used operationally for selection and/or classification at this time. However, future exploratory development concerned with potential modifications of the operational tests should, wherever feasible, include further evaluation of variables that were found to have significant validities in the present study.

2. Mental Group IVs should continue to be selected for naval service based on scores they obtain on the operational classification battery.

CONTENTS

	Page
INTRODUCTION	1
Background and Problem	1
Purpose	1
APPROACH	2
Experimental Test Development	2
Procedure	2
Predictors	3
Criteria	3
On-Job Performance	3
Ability to Progress into a Technical Rating	3
Length of Service and Attrition Characteristics	7
Analyses	7
RESULTS	9
Mental Group	9
Ratings to Which IVs were Assigned	9
Test-Retest Reliabilities of the Experimental Tests	12
Relative Performance of IVs and non-IVs	12
Predictors for On-job Performance of IVs and non-IVs	15
Potential Increases in Overall Validity for On-job Performance	18
Significant Predictors of IVs Able to Progress into Technical Ratings	22
Comparison of the Length of Service and Attrition Characteristics of Hi- and Lo-IVs	25
Racial Group	29
Predictors of On-job Performance of Blacks and Non-Blacks	29
Predictors for Rating Progression of Blacks and Non-Blacks	33
DISCUSSION AND CONCLUSIONS	41
Mental Level Group	41
Racial Group	43
RECOMMENDATIONS	45
REFERENCES	47
APPENDIX—DESCRIPTIONS OF THE 19 EXPERIMENIAL TESTS	A-0
DISTRIBUTION LIST	

LIST OF TABLES

		Page
1.	Predictors Derived from Experimental Tests	4
2.	Operational Predictors Used in the Study	6
3.	Incidence of Mental Group IVs in Navy Ratings	10
4.	Category IVs in Technical Ratings Classified by CREO Categories .	12
5.	Test-retest Reliabilities of the Cognitive Experimental Tests . .	13
6.	Supervisors' Evaluation of On-job Performance of IVs and non-IVs .	14
7.	Zero-order Validities of Operational Predictors for On-job Performance of IVs and Non-IVs	16
8.	Average Validity Coefficients of Operational Predictors for On-job Performance of IVs and Non-IVs	17
9.	Validities of Experimental Variables that were Statistically Significant for Either IVs or Non-IVs	19
10.	Predictors Providing Significant Increases to the Shrunken, Stepwise Multiple Rs for Job Performance of IVs	21
11.	Zero-order Validities of Operationally-derived Variables for Rating Progression of IVs	23
12.	Average Validity Coefficients of the Operationally-derived Variables for IVs for the A/T Criterion	24
13.	Experimental Variables Having Significant Zero-order Validities for IVs for the A/T Criterion	26
14.	Predictors Providing Significant Validity Increments to the Shrunken, Stepwise Multiple Rs of IVs for the A/T Criterion . . .	27
15.	Length of Service and Attrition Statistics for Hi- and Lo-IVs in the Total Sample	28
16.	Zero-order Validities of Operationally-derived Variables for On-job Performance of Blacks and Non-Blacks	39
17.	Average Validity Coefficients of the Operationally-derived Variables for On-job Performance of Blacks and Non-Blacks	31
18.	Significant Zero-order Validities of Experimental Variables for On-job Performance of Blacks and Non-Blacks	34
19.	Zero-order Validities of Operationally-derived Variables for the A/T Criteria for Blacks and non-Blacks	35

Page

20. Average Validity Coefficients of the Operationally-derived Variables for the A/T Criterion 36

21. Significant Zero-order Validities of Experimental Variables for Blacks and Non-Blacks for the A/T Criterion 38

22. Variables Providing Significant Validity Increments to the Shrunken Multiple Rs of Blacks and Non-Blacks for the A/T Criterion 40

INTRODUCTION

Background and Problem

Over the years, it has been widely believed that most personnel who score low on the Navy's entry aptitude measures tend to be of correspondingly low value to the service. This belief is supported by the correlations between scores obtained on these measures and success in technical training at Class "A" Schools. Generally, the lower the aptitude scores, the poorer the school performance, and the more limited the range of utilization. Consequently, personnel with lower aptitude scores generally have lower performance ratings, slower rates of advancement, and lower probabilities of being recommended for reenlistment. Since personnel at the lower end of the aptitude scale have been considered undesirable, one function of selection testing has been to reject them or to accept only as many as required to fill quotas.

Opponents of this policy have argued that men low on the aptitude scale could and should be used more effectively and in greater numbers. They believed that many low aptitude personnel possessed positive traits conducive to effective on-job performance, but that these traits were not being tapped by the entry tests because these tests had been developed to predict scholastic performance in Class "A" School. Thus, they maintained that, if appropriate tests were used (i.e., measures of attributes related to on-job performance and geared to the abilities of the low aptitude personnel), these personnel could be accepted in greater numbers and used more flexibly and profitably.

This philosophy was given concrete embodiment in the establishment of Project 100,000 in the late 1960s. Under this project, each service was required to accept prescribed, and unprecedentedly high, proportions of lower mental level personnel (e.g., personnel in the 10-30th centile on the AFQT). An important aspect of the project was the opportunity and the impetus that it gave for research into the selection, training, and utilization of low mental level personnel. More specifically, it provided a means for testing the hypothesis that this group could be usefully assessed by instruments tailored to its unique characteristics. (For summaries of research conducted under Project 100,000, see Ratliff and Earles, 1976.)

Purpose

The primary purpose of this effort, which originated as a consequence of Project 100,000, was to develop and validate tests that would be predictive of on-job performance of personnel who score low on traditional aptitude tests.

In addition, because of concern about possible bias against minorities, particularly Blacks, in the classification tests, a secondary purpose was to determine whether the operational classification tests were racially biased by computing and comparing validity coefficients for Blacks and non-Blacks. This aspect of the study required that the experimental tests be examined to determine whether they could replace any operational tests found to be biased.

APPROACH

Experimental Test Development

An extensive literature search was conducted to identify tests that had been useful for selecting low mental-ability personnel for either job or school assignments. A number of these tests or test types were mentioned by Ghiselli (1966).

Nineteen tests were selected for evaluation. These included eight measures of cognitive skills, four of motivation and perceptual/motor skills, three of vocational interests, and four of biographical/attitudinal data. These tests were modified for use in this study by eliminating as much of the reading material as possible, and by reducing the remaining material to the fifth grade reading level or below. In addition, the difficulty levels of the test items were relaxed so that the test means were about 70 percent of the maximum score. This was about 10 to 15 percentage points higher than the means of the operational tests. Descriptions of the modified tests are provided in the appendix.

Because of the relatively low difficulty levels of the experimental tests, it was expected that low mental ability personnel would react favorably to them. Thus, their test performance in relation to their abilities would be optimal. Also, because of the "experiential" nature of the experimental stimuli, the low reliance on reading skills, the relative absence of academic content from the test materials (particularly the absence of difficult or unusual vocabulary), it was hoped that the experimental tests would be "culture fair." By comparing the predictiveness of these experimental tests with that of the operational tests for the two major racial subgroups--Blacks and non-Blacks--it would be possible to determine whether tests in the experimental battery could be used to replace any operational tests to improve prediction of performance of Blacks.

Procedure

The experimental tests were divided into four experimental batteries, to be administered to four separate samples in four separate phases. In each phase, the testing was conducted early in recruit training, and supervisory follow-up ratings of on-job performance were obtained a few months before the conclusion of the enlistment period of the Mental Group IVs. For Phase 1, follow-up ratings were collected during the 21st and 46th months of enlistment. For Phases 2, 3, and 4, in which IVs had 2-year enlistment periods, the follow-up ratings were collected during the 21st and 22nd months.

Dates of test administration and collection of follow-up data are shown below:

Sample	Period during which Experimental Tests Were Administered	Date of Mailout for Supervisory On-job Performance Ratings
Phase 1	6/27/67 to 8/4/67	5/69 and 5/71
Phase 2	2/13/68 to 4/4/68	11/69
Phase 3	6/27/68 to 11/10/68	5/70
Phase 4	1/6/69 to 6/30/69	11/70

Samples of personnel tested during each phase ranged from 5,000 to 12,000 incoming recruits. For the follow-up ratings, from one-half to two-thirds of Mental Group I to III personnel were randomly eliminated for some phases to make the sample sizes of IVs and I-IIIs more nearly comparable. Follow-up sample sizes of Mental Group IVs ranged from 449 to 1890 personnel; and those of Mental Group I-IIIs, from 1650 to 3150.

Predictors

The experimental predictors are described in the appendix and listed, by study phase, in Table 1. The operational variables are presented in Table 2. The operational tests listed were subtests of the Basic Test Battery (BTB) that was used by the Navy at the time of data collection for personnel selection and classification. During the time that has elapsed since the last study phase, the BTB has been replaced by the Armed Services Vocational Aptitude Battery (ASVAB), which includes subtests designed to measure the same mental abilities as those in the BTB.

The last three variables listed in Table 2 were formed from binary codings of educational level, racial identification, and level of mental ability within the Mental Group IV range. For the last variable, personnel in the low IV range (AFQT scores 10 to 20, inclusive) were coded 0; and those in the high IV range (AFQT scores 21-30), 1.

Criteria

On-Job Performance

The mailout questionnaire on which supervisors rated personnel on-job performance included separate measures of global job performance covering the following points:

1. A measure of how much the supervisor would want or would not want to have the man reenlist.

2. A rating of the man's performance, using the categories of the operational enlisted performance evaluation form (NAVPERS 792).

3. A comparison of the man's overall performance to that of other personnel in the same rate.

Because responses to all three questions were similar, it was decided to use only responses to question 3 as the on-job performance criterion. Performances in the bottom 20 percent, lower 20-40 percent, middle 20 percent, upper 20-40 percent, and top 20 percent were coded 1, 2, 3, 4, and 5 respectively.

Ability to Progress into a Technical Rating

A binary 0-1 apprentice/technical (A/T) variable was formed to indicate whether subjects were assigned to apprenticeship or technical ratings at the time of the on-job performance follow-up. The apprenticeship group was defined as consisting of nonrated, nonstriker personnel in the Seaman, Fireman, Constructionman, or Airman ratings; and the technical group, of designated strikers and rated personnel. Personnel who were assigned to apprentice ratings were coded 0; and those who were assigned to technical ratings, 1. This A/T code was to be used to

Table 1

Predictors Derived from Experimental Tests

Experimental Test/Questionnaire	Experimental Predictor	Area of Measurement
	Phase I	
Memory for Numbers Test	Total Correct Recalls Weighted Total Correct Recalls First Error Span Score	Cognitive Skills
AFQT Retest	Vocab./Verbal Subtest Score Tool/Mech. Knowledge Subtest Score Arith. Reas. Subtest Score Spatial Reas. Subtest Score Total Retest Score	Cognitive Skills
Hand Skills Test	Part 3 Score Minus Base Rate Part 4 Score Minus Base Rate Total Score	Motivation and Perceptual/Motor Skills
Manual Speed Test	Part 3 Score Minus Base Rate Total Score	Motivation and Perceptual/Motor Skills
	Phase II	
Listening Skills Test I	Test Score	Cognitive Skills
Dominoes Test	Test Score	Cognitive Skills
Maze Test	Test Score	Motivation and Perceptual/Motor Skills
Job Check List	Clerical Scale Dangerous Military Scale Domestic Activities Scale Communications Scale Verbal Activities Scale Graphics Scale Clerical Hand Skills Scale	Vocational Interests
Biographical Information Form (BIF)	Father's Level of Education Father's Level of Occupation Mother's Level of Education Type of Neighborhood Both Parents in Home Relative Family Affluence Number of Siblings in College Number of Sibling High-school Dropouts Educational Attainment Expected Number of Books Read—Last 3 Mos. Luck vs. Hard Work for Success Problems in Getting Ahead Problems in Being Successful Socioeconomic Status Total BIF Retention Scale	Biographical/Attitudinal

Table 1 (Continued)

Experimental Test/Questionnaire	Experimental Predictor	Area of Measurement
	Phase III	
Listening Skills Test II	Test Score	Cognitive Skills
Strong Vocational Interest Blank (SVIB)	Business Management Scale Merchandising Scale Office Practices Scale Military Scale Technical Supervision Scale Mathematics Scale Science Scale Mechanical Scale Nature Scale Teaching Scale Navy Interest Scale Achievement Scale Liberal-Conservative Scale Masculinity-Femininity Scale Social Introversion Scale Occupational Level Scale	Vocational Interests
Performance Index (PI)	Personal Maturity Score General Maturity Score	Vocational Interests
Recruit Attitude Questionnaire (RAQ)	Fringe Benefits and Retirement Scale Equipment and Facilities Scale Patriotism and Duty Scale Travel and Adventure Scale Pay Scale Prestige Scale Advancement Scale	Biographical/Attitudinal
	Phase IV	
Card Pattern Test	Test Score	Cognitive Skills
Coding Test	Test Score	Cognitive Skills
Mechanical Principles Test	Test Score	Cognitive Skills
Word Finding Test	Test Score	Motivation and Perceptual/Motor Skills
Biographical Information Questionnaire (BIQ)	Factorially Derived Scales: Activities and Experiences Family Possessions Electro/Mechanical Interests Socioeconomic Status Social Leadership Logically Derived Scales: Impoverishment of Home Environment Teenage Activities Home Activities and Responsibilities Mechanical and Technical Experiences Recreation, Sports, and Hobbies Literary/Cultural Interests Travel Experiences Economic and Financial Responsibilities Interpersonal Interaction Socioeconomic Status Perceived Social Status	Biographical/Attitudinal
Recruit Temperament Survey (RTS)	Test Score	Biographical/Attitudinal

Table 2

Operational Predictors Used In the Study

Predictor	Acronym	Number of Items	Response Mode	Type of Scaling	Content Areas
Selection and Classification Tests					
Armed Forces Qualification Test Score	AFQT	100	Multiple Choice	Centile	Vocabulary, arithmetic reasoning, spatial reasoning, and mechanical knowledge.
General Classification Test Score	GCT	100	Multiple Choice	Navy Standard Score (N.S.S.)	Word meanings and the ability to reason verbally.
Arithmetic Reasoning Test Score	ARI	30	Multiple Choice	N.S.S.	Quantitative aptitude, including mathematical reasoning and problem solving.
Mechanical Test Score	MECH	100	Multiple Choice	N.S.S.	Basic mechanical and electrical knowledge and mechanical principles.
Clerical Test Score	CLER	50	Multiple Choice	N.S.S.	Perceptual speed and accuracy.
Sonar Pitch Memory Test Score	SONR	40	Multiple Choice	N.S.S.	Ability to perceive and remember small differences in tonal pitches.
Radio Code Aptitude Test Score	RADO	75	Multiple Choice	N.S.S.	Ability to learn, remember and use sound patterns as symbols.
Electronic Technician Selection Test Score	ETST	70	Multiple Choice	N.S.S.	Mathematics, science, electricity and electronics.
Shop Practices Score	SHOP	30	Multiple Choice	N.S.S.	Knowledge of tools and shop equipment.
Biographical Variables					
Year of Birth	YRBI		Free Answer	Integer	
Years of Education	YRED		Free Answer	Integer	
High School graduate versus non-High School graduate	HS		Binary	1-0	High School Graduate = 1 Non-High School Graduate = 0
Black versus non-Black	BL		Binary	1-0	Blacks Coded 1, non-Blacks coded 0.
Lo-IV versus Hi-IV	H4		Binary	1-0	Personnel with AFQT scores 10-20 were coded 0 (low-IVs), and those with AFQT scores 21-30 were coded 1 (Hi-IVs).

identify the distinguishing characteristics of personnel who progressed into technical ratings, since such ratings represent a high level of achievement for low-scoring men.

Length of Service and Attrition Characteristics

To permit analysis of the length of service and attrition characteristics of all Mental Group IVs in the sample, records were extracted from the Navy Enlisted History Tape (NEHT) (Naval Health Research Center, 1979). The NEHT contains longitudinal data covering the service history of all personnel who have enlisted in the Navy since 1965.

From the extracted data, two variables were selected for analysis: (1) the total number of months of enlisted service prior to discharge, and (2) the percentage of personnel who were categorized on the NEHT as having received attrition discharges. Attrition discharges are defined as terminations of enlisted service prior to the normal expiration of enlistment. Attrition terminations are discharges for such causes as disability, unsuitability, and misconduct.

Analysis

Evaluation of the responses to the four sets of experimental tests required from 2 to 4 years after testing for criterion information to mature. Project 100,000 was terminated in the meantime. The preliminary findings of this research were provided to the sponsors when they became available, and the analyses of the data were completed on a time-available basis.

When all data (test scores, biographical data, and supervisors' marks) had been collected, it was apparent that, despite the large samples, most ratings did not include enough personnel to compute reliable estimates of predictor-criterion correlation coefficients. Furthermore, it was not feasible, in terms of the time or resources available, to compute and publish rating-by-rating analyses of the predictor-criterion correlations. Therefore, it was decided to combine ratings into subgroups, and a number of subgrouping schemes were tried experimentally.

Correlational analyses were performed separately for personnel in the ratings with the largest Ns, for those in rating groups with similar duties, and for the total group. Statistics computed for apprenticeship and technical groupings had greater face validity and more consistency from phase to phase than did statistics based on the other categorization schemes. Therefore, the IV and non-IV personnel, subgrouped into apprenticeship and technical ratings, were systematically compared by means of an extensive series of analyses. The apprenticeship group was defined as consisting of nonrated, nonstriker personnel in the Seaman, Fireman, Constructionman, or Airman ratings; and the technical group, of designated strikers and rated personnel.

Small differences existed in the rating composition of the samples for the different phases. Stewards were eliminated from Phase 2 and from the first follow-up of Phase 1, as being an atypical group, largely Filipinos. The decision was reversed in later phases, however, when it was decided that information on

stewards would be of particular value. Tests of homogeneity of the predictor-criterion validity coefficients of the samples (described later) indicated that these minor differences in definition rules did not significantly alter the characteristics of the samples.

The ratings that included IV personnel were classified in terms of percentages of IVs assigned and of Career Reenlistment Objectives (CREO) characteristics. The mean overall on-job performance marks of IVs and non-IVs were compared. Zero-order validity coefficients between experimental and operational variables and the criteria were computed for each of the samples.

Multiple correlation coefficients were computed to provide an estimate of the potential of selected experimental variables for incrementing the validity of the operational classification battery. For this purpose, the following three batteries were formed and multiple correlations were computed for each, using step-wise multiple regression with an accretion paradigm:

1. The tests in the operational Navy classification battery.

2. The operational Navy Tests plus AFQT, YRED, HS, and YREI, which were readily available but were not being used for personnel classification.

3. All of the variables in (2) plus the experimental variables that had been administered to the sample.

The addition of variables to the selector battery was stopped at the point at which the first significant drop in the shrunken validity coefficient[1] occurred.

These analyses were carried out separately for the samples divided into IVs and non-IVs and subdivided by type of rating assignment (i.e., apprenticeship or technical). The data file was then divided into Blacks and non-Blacks and again subdivided by type of rating assignment. Separate analyses were carried out for category IVs using the criteria employed in the study: job performance, rating progression, and retention characteristics. For Blacks and non-Blacks, the analyses were carried out against on-job performance and rating progression.

---

[1]The shrunken multiple validity represents an estimate of the size of the validity coefficient that would be obtained from using the multiple regression weights derived on the original sample for a new sample of the same size from the same population. Thus, the shrunken coefficient is an estimate of the predictive accuracy achievable from the operational use of those variables. The following formula, recommended by Darlington (1968), was used to adjust the values of the sample mean square errors used to compute the shrunken coefficients.

$$\frac{N-2}{N-n-2} \cdot \frac{N+1}{N-n-1} S_{o(p)}^2$$

Where N = Number in the sample,

n = Total number of variables selected,

and

$S_{o(p)}^2$ = Sample mean square error.

RESULTS

This section summarizes the results of the experimental test development and validation effort. Although the data were collected over a decade ago, the findings are of current importance because many of the problems of Navy personnel operations have not changed substantially in the interim. In fact, some of the problems addressed herein have become even more critical during the years that have elapsed since some of the tests were originally administered. With the onset of the All-Volunteer military services and the consequent difficulties in filling enlistment quotas, the assignment of personnel to utilize their highest abilities has become even more important. In addition, the development and use of valid selection tests for minority and lower ability personnel continue to be of major concern to the Navy.

Mental Group

Ratings to which IVs were Assigned

Table 3 shows that, over the four phases, 18 percent of the IVs had been assigned to the 47 technical (T) ratings; and 82 percent, to the 4 apprenticeship (A) groups. Twenty-nine percent of the technical IVs were stewards (SDs); and 9.5 percent, equipment operators (EOs). The next nine ratings listed in Table 3—from BM through MM--contained between 3 and 5 percent of the technical IVs. Thus, 11 ratings, about a quarter of the total represented, contained 72 percent of the technical IVs.

For six of these 11 ratings--EO, BM, HM, YN, BU, and MM--more than half of the total IVs were concentrated in one or another of the phases. This suggests that special factors may have ballooned the incidence of IVs in a particular phase. For example, the fact that 68 of the 77 EOs were found in Phase 1 may be attributable to the fact that other research, which involved arbitrary assignment of groups of IVs to receive "A" School training for the EO rating (Standlee & Saylor, 1969) was being conducted at the time of data collection.

None of the 24 ratings in the bottom half of the distribution (from PC through AS) contained more than 1 percent of the total technical IVs. Further, for 10 of these ratings--AD, AE, FT, ET, AO, DT, CS, BU, AW, and AS, it appears that even this small percentage overrepresents the typical proportion of technical IVs. Based on an analysis of test-retest reliability of the Armed Forces Qualification Test completed during Phase 1, it was concluded that all Phase 1 IVs found in these 10 ratings were probably I-IIIs who had been spuriously classified as IVs. As shown in Table 3, IVs were found in the AW and AS ratings only during Phase 1. Further, in the FT, AE, DT, and ET ratings, the IVs in Phase 1 made up 50 percent or more of the total IVs in those ratings in the study. Therefore, the 18 percent of technical IVs shown in Table 3 is probably a slight overestimation of the percentage of IVs that would be expected to advance into technical ratings under normal conditions.

The total numbers and percentages of technical IVs in the Career Reenlistment Objectives (CREO) categories are shown in Table 4. About 21 percent were serving in ratings having shortages in retention; and about 15 percent, in ratings that already had an oversupply of career personnel.

Table 3

Incidence of Mental Group IVs in Navy Ratings

Rating	Abbrev	Male s							Return							CREO Category[a]
		Phase					% of total Tech. IVs	% of Total IVs	Phase					% of total Tech. IVs	% of total IVs	
		1	2	3	4	Total			1	2	3	4	Total			
Technical																
Steward	SD	37	1[b]	111	87	236	29.0	5.2	-	-	49	29	78	20.5	3.8	C
Equipment Operator	EO	1	3	68	5	77	9.5	1.7	1	-	11	-	12	3.2	.6	D
Boatswain's Mate	BM	31	6	-	1	38	4.7	.8	27	3	-	1	31	8.1	1.5	C
Hull Technician	HT	13	4	10	9	36	4.4	.8	9	2	4	3	18	4.7	.9	B
Hospital Corpsman	HM	3	7	6	19	35	4.3	.8	2	6	3	10	21	5.5	1.0	C
Commissaryman	CS	13	5	8	8	34	4.2	.8	11	3	4	3	21	5.5	1.0	C
Yeoman	YN	6	2	7	15	27	3.3	.6	4	1	4	9	18	4.7	.9	C
Builder	BU	1	1	19	6	27	3.3	.6	1	1	1	1	4	1.0	.2	C
Radioman	RM	10	3	7	7	27	3.3	.6	7	1	6	6	20	5.2	1.0	C
Engineman	EN	12	6	2	4	24	3.0	.5	11	2	-	1	14	3.7	.7	C
Machinist's Mate	MM	19	2	2	1	24	3.0	.5	15	2	1	1	19	5.0	.9	B
Storekeeper	SK	8	3	3	7	21	2.6	.5	4	2	1	3	10	2.6	.5	C
Boiler Technician	BT	10	4	2	2	18	2.2	.4	7	1	1	1	10	2.6	.5	B
Utilities Man	UT	1	-	16	1	18	2.2	.4	1	-	3	-	4	1.0	.2	C
Construction Mechanic	CM	2	2	11	3	18	2.2	.4	1	1	4	1	7	1.8	.3	D
Ship's Serviceman	SH	5	6	5	1	17	2.1	.4	4	3	4	-	11	2.9	.5	B
Signalman	SM	1	2	2	11	16	2.0	.4	1	1	-	5	7	1.8	.3	B
Electrician's Mate	EM	4	6	-	2	12	1.5	.3	4	2	-	2	8	2.1	.4	C
Aviation Machinist's Mate	AD	3	1	2	5	11	1.3	.3	1	1	2	3	7	1.8	.3	B
Quartermaster	QM	4	4	-	2	10	1.2	.2	2	3	-	1	6	1.6	.3	B
Aviation Structural Mechanic	AM	2	-	-	7	9	1.1	.2	2	-	-	7	7	1.8	.3	B
Construction Electrician	CE	-	-	7	2	9	1.1	.2	-	-	2	-	2	.5	.1	F
Personnelman	PN	5	1	-	3	9	1.1	.2	4	-	-	1	5	1.3	.3	C
Postal Clerk	PC	3	-	-	3	6	.7	.1	3	-	-	3	6	1.6	.3	F
Communication Technician	CT	-	3	1	2	6	.7	.1	-	3	1	2	6	1.6	.3	C
Operations Specialist	OS	2	-	1	2	5	.6	.1	2	-	1	1	4	1.0	.2	A
Machinery Repairman	MR	-	2	-	3	5	.6	.1	-	2	-	3	5	1.3	.3	A
Disbursing Clerk	DK	1	1	1	1	4	.5	.1	1	1	1	-	3	.8	.2	D

[a]Category A--Extreme shortage of personnel in rate.
Category B--Shortage of personnel in rate.
Category C--Rate manning is approximately correct.
Category D--Rate manning is excessive; voluntary conversions are recommended.
Category E--Rate manning is excessive; involuntary conversions may be directed.

[b]Although it was decided to eliminate stewards (SDs) from Phase 2 (see p. 7), one SD took the test in error.

Table 3 (Continued)

Rating	Abbrev	Mailout Phase 1	2	3	4	Total	% of Total Tech. IVs	% of Total IVs	Return Phase 1	2	3	4	Total	% of Total Tech. IVs	% of Total IVs	CRFO Category*
Technical (Continued)																
Fire Control Technician	FT	3	-	-	1	4	.5	.1	-	-	-	-	0	0.0	.0	D
Gunner's Mate	GM	1	1	-	1	3	.4	.1	1	-	-	-	1	.3	.1	B
Aviation Electrician's Mate	AE	2	-	-	1	3	.4	.1	1	-	-	1	2	.5	.1	B
Aviation Ordnanceman	AO	1	-	2	-	3	.4	.1	1	-	-	-	1	.3	.1	B
Steelworker	SW	-	-	3	-	3	.4	.1	-	-	1	-	1	.3	.1	C
Torpedoman's Mate	TM	-	1	1	-	2	.3	.1	-	1	1	-	2	.5	.1	B
Dental Technician	DT	1	-	-	1	2	.3	.1	1	-	-	1	2	.5	.1	E
Electronics Technician	ET	1	1	-	-	2	.3	.0	-	-	-	-	0	0.0	.0	C
Mineman	MN	1	-	1	-	2	.3	.0	-	-	1	-	1	.3	.1	B
Aviation Boatswain's Mate	AB	1	-	-	-	1	.1	.0	1	-	-	-	1	.3	.0	B
Aerographer's Mate	AG	-	-	-	1	1	.1	.0	-	-	-	-	0	0.0	.0	C
Aviation Maintenance Administrationman	AZ	-	-	-	1	1	.1	.0	-	-	-	1	1	.3	.0	C
Sonar Technician	ST	-	-	1	-	1	.1	.0	-	-	-	-	0	0.0	.0	C
Aviation Storekeeper	AK	1	-	-	-	1	.1	.0	1	-	-	-	1	.3	.0	E
Instrumentman	IM	-	1	-	-	1	.1	.0	-	1	-	-	1	.3	.0	B
Molder	ML	1	-	-	-	1	.1	.0	1	-	-	-	1	.3	.0	B
Aviation electronics Technician	AT	1	-	-	-	1	.1	.0	-	-	-	-	0	0.0	.0	B
Aviation Anti-Submarine Warfare Operator	AW	1	-	-	-	1	.1	.0	1	-	-	-	1	.3	.0	B
Aviation Support Equipment Technician	AS	1	-	-	-	1	.1	.0	1	-	-	-	1	.3	.0	C
Total		213	79	296	225	813	100.0	18.0	134	43	106	98	381	100.0	18.5	
Apprenticeship																
Seaman	SN	141	355	540	1128	2164	-	47.8	97	191	365	416	969	-	47.0	
Fireman	FN	71	368	242	405	1091	-	24.1	53	189	214	134	490	-	23.8	
Airman	AN	21	66	151	125	363	-	8.0	15	38	82	65	200	-	9.7	
Construction man	CN	3	49	36	7	95	-	2.1	3	12	4	1	20	-	1.0	
Total		236	838	974	1665	3713	-	82.0	168	430	465	616	1679	-	81.5	
TOTAL		449	917	1270	1890	4526	-	100.0	302	473	371	714	2080	-	100.0	

*Category A--Extreme shortage of personnel in rate.
Category B--Shortage of personnel in rate.
Category C--Rate manning is approximately correct.
Category D--Rate manning is excessive; voluntary conversions are recommended.
Category E--Rate manning is excessive; involuntary conversions may be directed.

Table 4

Category IVs in Technical Ratings Classified
by CREO Categories

CREO Category[a]	Mailout		Return	
	N	%	N	%
A	10	1.2	9	2.3
B	159	19.6	96	25.2
C	523	64.3	243	63.8
D	103	12.7	22	5.8
E	18	2.2	11	2.9
Total	813	100.0	381	100.0

[a]Category A—Extreme shortage of personnel in rate.
Category B—Shortage of personnel in rate.
Category C—Rate manning is approximately correct.
Category D—Rate manning is excessive; voluntary conversions are recommended.
Category E—Race manning is excessive; involuntary conversions may be directed.

Test-Retest Reliabilities of the Experimental Tests

Test-retest reliabilities of the major experimental tests measuring cognitive skills or motivation and perceptual/motor skills were computed on separate full-range samples of recruits at NTC San Diego. For this part of the study, the tests were administered in June 1971; and the retests, about 5 weeks later. As shown in Table 5, most of the experimental tests had reliabilities in the 60s and 70s. For the Memory for Numbers and the Card Pattern tests, however, reliabilities were in the .30s—a level that would have to be improved before the tests could be used operationally.

The major practical consequence of the relatively low reliabilities of the tests in the experimental batteries would be to lower the observed validities for these tests. This effect slightly penalized the validity coefficients obtained for the experimental tests in relation to those obtained for the operational tests, which had test-retest reliabilities in the .70s and .80s.[2]

Relative Performance of IVs and non-IVs

Table 6 presents the mean supervisory evaluations of non-IVs and IVs and the correlation coefficient of H4 with those evaluations. Since H4 is a binary variable in which Hi-IVs were coded "1" and Lo-IVs were coded "0," r_{H4} is a point biserial correlation coefficient whose sign indicates which of the mental level groups had the higher average supervisory mark. As shown, the mean on-job performance evaluation of IVs was lower than that of non-IVs for all of the samples

[2]Swanson, L. Personal communication, 1977.

Table 5

Test-retest Reliabilities of the Cognitive Experimental Tests

Predictor	Phase Admin.	r_{tt}	Sample N
Cognitive Skills Test			
Memory for Numbers, Total Correct	1	.34	110
Listening Skills Test II Score	3	.69	129
Dominoes Test Score	2	.66	172
Card Pattern Test Score	4	.33	177
Coding Test Score	4	.47	141
Motivation and Perceptual/Motor Skills			
Hand Skills Test:	1		
Part 3 score minus base rate		.67	172
Part 4 score minus base rate		.71	172
Total Score		.68	172
Manual Speed Test Score	1	.57	133
Maze Test Score	2	.77	149
Word Finding Test Score	4	.66	143

Table 6

Supervisors' Evaluation of On-job Performance of IVs and Non-IVs

Sample	N		Mean			r_{H4}
	Non-IV	IV	Non-IV	IV	Non-IV-IV	
Apprenticeship Group						
Phase 1						
1st Follow-up	21	168	2.57	3.11	-.54*	-.04
2nd Follow-up	346	90	3.68	3.32	.36**	.19
Phase 2	888	431	3.31	2.94	.37**	.05
Phase 3	628	465	3.48	2.92	.56**	.08
Phase 4	597	616	3.56	3.20	.36**	.08
Technical Group						
Phase 1						
1st Follow-up	271	134	3.73	3.58	.15	.01
2nd Follow-up	1104	133	3.76	3.41	.35**	.01
Phase 2	2686	43	3.69	3.67	.02	.26
Phase 3	1718	105	3.75	3.36	.39**	.19
Phase 4	1629	98	3.77	3.32	.45**	.09

*p < .05.
**p < .01.

except the apprenticeship group sample in the first follow-up of Phase 1. The small size of this sample, however, suggests that its mean was very unreliable, particularly since the mean evaluation of non-IV personnel in the apprenticeship group in the second sample for Phase 1 corresponds with those of their counterparts in the other phases. The mean evaluation of non-IVs was significantly higher than that of IVs in four of the apprenticeship group samples and in three of the technical group samples.

Although the mean decrements of IVs for on-job performance ranged from .02 to .56, for most samples they were within .35 to .45. Since the standard deviations of the job performance marks for non-IVs in these groups (not shown) generally ranged between 1.00 and 1.20, decrements of this magnitude would indicate that the averages for IVs were at about the 33rd to 39th percentile levels on the non-IVs distribution for on-job performance (compared with an average performance at the 50th percentile level).

Predictors for On-job Performance of IVs and non-IVs

1. Operational

Validities of the operational variables for on-job performance are shown in Table 7. Because the major objective of the study was to identify any variable that is predictive of the on-job performance of IVs only, no correction for restriction of range was made.

The 93 significant validities shown in Table 7 range in magnitude from .05 to .44. All but one of the 65 significant validities of the operational tests are positive, indicating that high scores on the tests were generally associated with high on-job performance. Thus, although the operational classification tests had been developed to predict "A" School grades, it appears that they are also useful for predicting the on-job performance of non-"A" School personnel.

Because the definitions of the variables shown in Table 7 did not change from one sample to another, it was possible to average the validity coefficients across samples to obtain a more stable estimate of the population value. This coefficient would more accurately describe the ability of a variable to predict on-job performance than would any single coefficient. Therefore, Fisher's z transformations of the validity coefficients for the first follow-up of Phase 1 and for Phases 2, 3, and 4 (the counterpart of the Phase 1 1st follow-up in terms of elapsed service) were prepared and tested for significance using the $\underline{X}^2$ statistic. For the variables for which the $\underline{X}^2$ value was not significant (indicating that the sample values were statistically equivalent), weighted average coefficients were computed. For those variables for which the $\underline{X}^2$ test was significant, the coefficients for the individual samples were examined. If two or three of the coefficients were similar in magnitude, but the other(s) differed considerably, the outlier(s) was excluded and the $\underline{X}^2$ coefficient recomputed on the remaining coefficients. Weighted means were computed for any of the variables for which the $\underline{X}^2$ values at this stage were not significant. As a result of this second evaluation, weighted means were computed for all of the remaining variables, as shown in Table 8. The averages that are based on fewer than four coefficients are identified by superscripts.

Table 7

Zero-order Validities of Operational Predictors for On-job Performance of IVs and Non-IVs

Predictor Variable	Phase 1								Phase 2				Phase 3				Phase 4			
	1st Follow-up				2nd Follow-up															
	IV		N-IV		IV		N-IV		IV		N-IV		IV		N-IV		IV		N-IV	
	r	N	r	N	r	N	r	N	r	N	r	N	r	N	r	N	r	N	r	N
Apprenticeship Group																				
AFQT	-.03	167	.04	21	.23*	87	.03	346	.07	431	.05	889	.08	474	.08	636	.09*	615	.13*	597
GCT	-.15*	168	.08	21	.13	90	.09	345	.09*	431	.06	889	.08	471	.09*	635	.06	608	.11*	594
ARI	-.11	168	.44*	21	.05	90	05	345	.14**	431	.10**	889	.14**	471	.11**	635	.08	608	.14**	594
MECH	.14	168	-.34	21	-.01	90	.07	345	.13**	431	.07*	889	.12*	470	.08	635	.12**	608	.13*	594
CLER	.03	167	.02	21	-.06	89	.10	343	.11*	431	.10**	889	.14**	471	.10*	635	.16**	608	.08	594
SONR	-.08	168	-.15	21	.06	85	.09	343	.05	431	.00	889	.12*	469	.01	634	.04	608	.10*	594
RADO	.10	145	.31	20	-.07	73	.00	338	.12*	431	.06	889	.12*	394	.01	625	.14**	541	.12**	594
ETST	.09	168	.28	21	.04	85	.11*	343	.00	431	.05	889	-.06	469	.10*	634	.00	608	.06	594
SHOP	.01	168	-.35	21	.00	89	.00	345	.13**	431	.06	889	.06	471	.09*	635	.10*	608	.10*	594
Year of Birth (YRBI)	-.04	168	-.11	21	.10	90	-.11*	346	-.11*	431	.02	889	---	---	---	---	.12**	616	.07	597
High School Graduate (HS)	.03	168	.43*	21	.04	90	.06	346	-.13	431	.10**	889	.10*	474	.07	636	.10*	616	.07	597
Black/non-Black (BL)	.02	168	.00	21	.07	90	.05	346	.00	413	-.05	779	-.14**	474	.05	636	-.03	616	.0	597
Years of Education (YRED)	-.02	168	.35	21	.07	90	.16**	346	.32	20	.11**	854	.07	474	.18**	636	.09*	616	.10*	597
Lo-IV/Hi-IV (H4)	-.04	168	---	---	.19	90	---	---	.05	431	---	---	.08	474	---	---	.08	616	---	---
Technical Group																				
AFQT	.09	134	.11*	271	-.05	122	.06*	1104	.23	43	.04	2686	.26**	105	.07*	1731	.09	98	.08**	1629
GCT	.01	131	.04	271	-.05	131	.07*	1098	.07	43	.07**	2686	.09	104	.05	1721	.08	98	.12**	1613
ARI	.08	131	.09	271	-.02	131	.12**	1098	.17	43	.07**	2686	.08	104	.11**	1721	.14	98	.13**	1613
MECH	-.02	131	.06	271	.06	131	-.01	1098	.11	43	.05*	2686	-.03	103	.08**	1721	.09	98	.09**	1613
CLER	.04	131	.05	271	.10	130	.18**	1092	-.06	43	.09**	2686	.01	104	.07*	1721	.29**	98	.12**	1613
SONR	.12	131	-.03	271	-.16	93	.01	1067	-.18	43	.03	2686	.09	73	.03	1659	.03	98	-.02	1612
RADO	-.01	124	.05	270	-.09	90	.07*	1054	-.14	43	.04	2686	-.02	62	.08**	1646	.11	94	.04	1595
ETST	.07	131	.09	271	.04	93	.09**	1067	.01	43	.07**	2686	.23*	73	.12**	1659	.17	98	.08**	1612
SHOP	-.02	131	.01	271	.10	128	.00	1097	.13	43	.05*	2686	.01	104	.07*	1721	.06	98	.11**	1612
Year of Birth (YRBI)	-.02	134	-.07	271	-.16	133	-.10**	110	.06	43	-.09**	268	---	---	---	---	-.30**	98	-.14**	[illegible]
High School Graduate (HS)	.08	134	.03	271	.16	133	.07*	1104	-.22	43	.07**	2685	.33**	105	.03	1731	.36**	98	.18**	1629
Black/non-Black (BL)	.15	134	-.06	271	-.16	133	.01	1104	.02	42	-.01	2420	.12	105	.02	1731	-.03	98	-.02	1629
Years of Education (YRED)	.16	134	.01	271	.02	133	.12**	1104	-.49	9	.12**	2636	.29**	105	.14**	1731	.35**	98	.17**	1629
Lo-IV/HI-IV (H4)	.01	134	---	---	.01	133	---	---	.26	43	---	---	.19	105	---	---	.09	98	.00	1629

*p < .05.
**p < .01.

Table 8

Average Validity Coefficients of Operational Predictors for On-job Performance of IVs and Non-IVs

Operational Predictor	IVs		Non-IVs	
	Wtd. Mean r	N	Wtd. Mean r	N
Apprenticeship Group				
AFQT	.07*	1687	08*	2143
GCT	.07*	1510[a]	.08*	2139
ARI	.09**	1678	.12**	2139
MECH	.12**	1677	.09**	2139
CLER	.13**	1677	.09**	2139
SONR	.05	1676	.03	2138
RADO	.13**	1511	.06*	2128
ETST	-.01	1676	.07*	2138
SHOP	.09**	1678	.08*	2139
YRBI	-.09*	599[b]	.04	1507[a]
HS	.10**	1090[b]	.09**	2143
BL	-.05	1671	.02	2033
YRED	.07*	1278	.13**	2108
H4	.07*	1689	--	----
Technical Group				
AFQT	.15**	380	.06	6317
GCT	.06	376	.08*	6291
ARI	.11*	376	.10**	6291
MECH	.02	375	.07*	6291
CLER	.09	376	.09**	6291
SONR	.05	345	.01	6228
RADO	.01	323	.05	6197
ETST	.12*	345	.09**	6228
SHOP	.02	376	.07*	6290
YRBI	-.11*	275[a]	-.11**	4583[a]
HS	.25**	337[a]	.09**	6316[c]
BL	.10*	379	-.01	6051
YRED	.30**	212[a]	.14**	6267
H4	.11*	380	--	----

[a]Wtd. mean based on data from only three samples.

[b]Wtd. mean based on data from only two samples.

[c]Mean was computed from nonhomogeneous rs ($\underline{X}^2 = 21.89$, $\underline{p} < .01$).

*$\underline{p} < .05$.

**$\underline{p} < .01$.

In the apprenticeship group, average validity coefficients showed statistically significant prediction of on-job performance of IVs for seven of the nine operational tests, with RADO, MECH, and CLER having the highest validities. In addition, both education variables (HS and YRED) and the year of birth (YRBI) were statistically significant.

In the technical group, three operational tests--AFQT, ARI, and ETST--and all five biographical variables were significant predictors of on-job performance of IVs. HS and YRED, the two biographical variables with the highest coefficients, were more predictive of on-job performance of IVs than any of the operational tests. The coefficients for the other biographical variables indicate that the IVs who performed best in technical ratings tended to be older, Black, and Hi-IV.

Operational tests were also significant predictors of on-job performance of non-IVs in both job groups, although the correlations for non-IVs tended to be smaller than those for IVs.

2. Experimental

Table 9 presents validity coefficients of the experimental variables for the apprenticeship and technical groups. As shown, for IVs in the apprenticeship group, scores of two cognitive skill measures (Coding and Word Finding), responses to two Biographical Information Form (BIF) items, and three scales from the Biographical Information Questionnaire (BIQ) had significant validities. The better performing apprenticeship personnel had the following characteristics:

a. Relatively high abilities in coding and word finding.
b. Self-reported admission that he has little chance in life.
c. Raised in a two-parent home.
d. Self-reported leadership and interpersonal relations skills.
e. Self-reported interests and past experience in mechanical types of activities.

Sixteen experimental predictors had statistically significant validities for IVs in the technical group. They included scores on (a) four cognitive tests--AFQT Vocab./Verbal, Memory for Numbers, Coding, and Mechanical Principles, (b) five measures of occupational interests (SVIB scales), (c) six measures of personal interests and preferences, and (d) a measure of neuroticism. The better technical IVs tended to (a) have high scores on the four cognitive tests listed, (b) have little interest in graphics and routine clerical activities, and (c) be highly interested in office practices, mathematics, science, teaching, and achievement. In addition, they were more interested in electromechanical activities, had a wider and more diversified social and educational background, and had more balanced, less neurotic personalities than other IVs.

Potential Increases in Overall Validity for On-job Performance

The next step consisted of identifying the most effective subsets of predictors within three sets of variables:

Table 9

Validities of Experimental Variables that were Statistically Significant for Either IVs or Non-IVs

	Validity for Global Performance Marks							
	Apprenticeship Group				Technical Group			
	IV		Non-IV		IV		Non-IV	
Experimental Predictor	r	N	r	N	r	N	r	N
Phase 1								
AFQT Retest								
Vocab./Verbal Subtest Score	—	—	—	—	.16*	134	-.05	270
Memory for Numbers Test								
Weighted Total Correct	—	—	—	—	.18*	132	.07	271
Phase 2								
Dominoes Test Score	—	—	—	—	.04	19	.08**	1191
Biog. Info. Form (BIF) Items								
Father's level of education	—	—	—	—	.41	22	-.09**	1788
Both parents in home	.15*	260	.11*	569	—	—	—	—
Number of siblings who went to college	—	—	—	—	-.32	26	.07**	1771
Level of educational attainment expected of individual by parents	—	—	—	—	.04	26	.09**	1787
Opinion that individual has little chance of success in life	.13*	220	.00	525	.09	26	.06*	1826
Retention scale	.11	424	.10*	781	.01	43	.08**	2461
Job Check List Scales								
Clerical	-.06	349	-.10*	779	—	—	—	—
Domestic activities	—	—	—	—	-.03	36	.06**	2461
Graphics	—	—	—	—	-.39*	36	-.01	2461
Clerical hand skills	—	—	—	—	-.37*	36	-.05*	2461
Phase 3								
RAQ Scales								
Fringe benefits and retirement	—	—	—	—	.47	6	.23**	274
Equipment and facilities	.02	57	.20**	169	—	—	—	—
Travel and adventure	—	—	—	—	.93**	6	-.10	274
Pay	—	—	—	—	-.07	6	-.19**	274
Advancement	.18	57	-.22**	169	—	—	—	—
SVIB Scales								
Business management	—	—	—	—	.20	73	.07*	1698
Office practices	—	—	—	—	.25*	73	.08**	1698
Merchandising	—	—	—	—	.20	73	.07*	1698
Technical supervision	—	—	—	—	.22	73	.10**	1698
Mathematics	—	—	—	—	.33**	73	.08**	1698
Science	—	—	—	—	.23*	73	.03	1698
Mechanical	—	—	—	—	.18	73	.06*	1698
Teaching	—	—	—	—	.28*	73	-.01	1698
Navy interest	—	—	—	—	.14	73	.08**	1698
Achievement	—	—	—	—	.24*	73	.00	1698
Liberal/conservative	—	—	—	—	-.14	73	-.10**	1698
Masculinity/femininity	—	—	—	—	.01	73	.09**	1698
Phase 4								
Coding Test Score	.14**	615	.11**	597	.28**	98	.11**	1629
Word Finding Test Score	.15*	222	.13*	356	.14	98	.09**	1629
Mechanical Principles Test Score	—	—	—	—	.50**	24	.05	740
BIQ Scales								
Logically Derived:								
Mechanical and technical experience	.14*	308	-.08	412	—	—	—	—
Economics and financial responsibility	.07	308	-.07	412	.33*	35	.04	960
Literary/cultural interest	.07	308	-.13**	412	—	—	—	—
Interpersonal interaction	.16**	308	-.05	412	.06	35	.07*	960
Perceived social status	.04	308	.13**	412	—	—	—	—
Factorially Derived:								
Activities and experiences	—	—	—	—	.50**	32	.03	909
Electro/mechanical interest	.02	236	.27**	359	.38**	32	-.07*	909
Socioeconomic status	—	—	—	—	.15	32	.03*	909
Social leadership	.15*	236	-.06	359	.19	32	-.09*	909
Recruit Temperament Survey (RTS) Score	-.11	290	-.16*	172	-.32**	61	-.05	639

*$p < .05$.
**$p < .01$.

1. The operational Navy aptitude tests (Basic Test Battery (BTB) and Special Tests) used to select for Class "A" Schools.

2. The operational aptitude tests plus other variables of operational record not normally used to select for Class "A" Schools.

3. The whole set of operational and experimental variables.

The variables selected for these three predictor batteries are shown in Table 10. The variable selected at each step is shown, together with the shrunken R obtained for the complete battery selected up to and including that step.

Table 10 makes explicit some relationships that are implied in Tables 8 and 9: The experimental tests were not very useful supplements to operational variables for predicting on-job performance of Mental Group IVs. For IVs in the apprenticeship group, although considerable variation occurred from sample to sample, the data in Table 10 suggest that an optimally weighted subset of operational tests will predict on-job performance 20 to 22 months later at a shrunken validity between .10 and .20. Biographical variables added to the predictor set will improve this coefficient about .04 or .05. Based on the findings from Phase 4, selected scales for personality interests and neuroticism may increase this coefficient 6 to 13 points. None of the other experimental variables would add to the predictive ability already present in operational tests and biographical variables. Thus, the maximum predictability of on-job performance for IVs in apprenticeship ratings over a 2-year interval would be about .40.

Experimental tests were more useful for predicting on-job performance of IVs in technical than in apprenticeship ratings. These tests incremented the maximum predictability of the operational variables for all four samples of the technical group; for the Phase 1 and 2 samples, experimental variables were more predictive than any of the operational variables. Except for the weighted score on the Memory for Numbers Test, the experimental variables that significantly added predictiveness to the operational variables were measures of personal interests and/or activities.

The sample to sample differences in the variables that were selected for multiple regression of IVs in the technical group and in the maximum accuracy of prediction were so great that no firm conclusions concerning the predictiveness of the operational variables can be drawn. Although the multiple regression validity characteristics of operational tests from sample to sample are inconsistent, the validity coefficient for HS is high and statistically significant for two of the technical samples (Phases 3 and 4). The fact that these two phases include the largest samples suggests that HS may be consistently predictive for IVs in technical ratings. These considerations suggest that the maximum shrunken validities of the operational variables (subset 2) for IVs in the technical group are in the high 20s. Measures of interests, background, and experience--such as are obtained by the Strong Vocational Interest Blank and by empirically-derived biographical information scales--could increase this coefficient considerably. The maximum shrunken validity of an optimal battery of operational and experimental variables for on-job performance of IVs in technical ratings over a 2-year interval might be expected to be in the .30 to .40 range, but the small samples make this problematical.

Table 10

Predictors Providing Significant Increases to the Shrunken, Stepwise Multiple Rs for Job Performance of IVs

Set of Predictors	Phase 1		Phase 2			Phase 3			Phase 4		
	Shrunken R	Variable Selected	Shrunken R	Variable Selected	N	Shrunken R	Variable Selected	N	Shrunken R	Variable Selected	N
			Apprenticeship Group								
1. Operational Navy Classification Tests		None	.10 .14	ARI MECH	430 430	.10 .15	CLEP ARI	465 465	.15 .18 .21	CLER MECH RADO	616 616 616
2. Variables in set (1) plus AFQT, YRBI, YRED, and HS		None	.10 .14 .17	ARI MECH HS	430 430 430	.10 .15	CLER ARI	465 465	.15 .20 .23 .26	CLER YRBI MECH RADO	616 616 616 616
3. Variables in set (2) plus the experimental tests		None	.10 .14 .17	ARI MECH HS	430 430 430	.10 .15	CLER ARI	465 465	.15 .20 .23 .26 .29 .39	CLER YRBI MECH RADO Interpersonal Interaction RTS	616 616 616 616 412 172
			Technical Group								
1. Operational Navy Classification Tests		None		None		.01	ETST	73	.21	CLER	98
2. Variables in set (1) plus AFQT, YRBI, YRED, and HS		None		None		.27	HS	105	.30 .36 .39	HS YRBI CLER	98 98 98
3. Variables in set (2) plus the experimental tests	.03	Memory for Numbers, Weighted Total 134	.23	Graphics Interest	35	.27 .32	HS SVIB Math Scale	105 73	.30 .53 .71	HS Mechanical Principles Activities & Experiences	98 24 24

The substantial sample-to-sample differences in the operational variables that were selected for multiple regression of IVs in the technical group and in the magnitudes of the validity coefficients appear to be inconsistent with the previous findings of no statistically significant intersample differences in the validity coefficients of the operational tests. These discrepencies probably occurred because (1) the scores of the operational tests were highly intercorrelated, and (2) there were substantial differences in sample sizes (some were quite small). This resulted in relatively large permissible ranges of variation in the validity coefficients before a statistically significant difference would be found.

The variation in the size of the validity coefficients of operational tests would result in differences among samples in the variable that was initially selected for multiple regression. The high intercorrelations among the variables would then prevent the operational tests that were most similar to the tests initially selected from being selected on subsequent steps. Given these conditions, relatively small sample-to-sample variation in magnitudes of the validity coefficients could result in predictive composites for the operational variables, which were composed of few if any of the same variables from one sample to another. In addition, the fact that shrunken coefficients were used for the multiple regression statistics, wheras the homogeneity comparisions were carried out on the zero-order coefficients would make a difference. Nevertheless, the fact remains that the only sample-to-sample consistency in the predictive relationships among the operational variables for IVs in technical ratings as determined by multiple regression was for HS. Conclusions concerning this predictive relationship should be considered tentative until they have been verified by additional research.

Significant Predictors of IVs Able to Progress into Technical Ratings

Mental Group IVs who are capable of progressing into technical ratings have more potential as career enlisted personnel in the Navy than do IVs who are capable of performing in apprenticeship ratings but not of advancing into a technical rating. Thus, to determine the characteristics of IVs who progress into technical ratings, the zero-order validities of the predictor variables with the apprenticeship/technical (A/T) code or criterion were computed.

Table 11 presents these validity coefficients for the operationally-derived variables for the IVs in Phases 1 through 4. As shown, 42 of the 55 coefficients were statistically significant. The only variables that did not consistently have significant validity coefficients from one sample to another were MECH, HS, and the binary Black/non-Black variable.

Averages for the coefficients in Table 11 are shown in Table 12. Among the Navy classification tests, GCT, ARI, and ETST were the most effective predictors of rating progression of IVs. Education and biographical variables, which were the most effective predictors of on-job performance, were only mediocre for predicting rating progression of IVs in their first 22 months of service. The data in Table 12 must be interpreted cautiously because of the extreme variability in the coefficients averaged.

Table 11

Zero-order Validities of Operationally-derived Variables for Rating Progression of IVs

Predictor Variables	Phase 1		Phase 2		Phase 3		Phase 4	
	r	N	r	N	r	N	r	N
AFQT	.22**	435	.11**	972	.15**	579	.14**	1889
GCT	.25**	445	.23**	972	.06	575	.26**	1876
ARI	.24**	445	.27**	972	.09*	575	.17**	1876
MECH	.07	445	.00	972	-.03	575	.07*	1876
CLER	.23**	442	.16**	972	.12**	573	.12**	1876
SONR	.14**	401	.07*	972	.02	542	.08**	1875
RADO	.20**	366	.16**	972	.15**	456	.08**	1688
ETST	.18**	401	.21**	972	.06	542	.12**	1875
SHOP	.20**	441	.10**	972	-.03	575	.12**	1876
YEAR OF BIRTH (YRBI)[a]	-.14**	449	-.10**	971	--	---	.06*	1890
HIGH SCHOOL (HS)	.24**	449	.12**	972	.06	579	.04	1890
BLACK/NON BL. (BL)	.02	449	.06	935	-.13**	579	.11**	1890
YEARS OF EDUC. (YRED)	.25**	449	.43**	43	.03	579	.07*	1890
LO-IV/HI-IV (H4)	-.18**	449	.09**	972	.15**	579	.11**	1890

[a]YRBI data were not collected for the Phase 3 sample.

*$p < .05$.

**$p < .01$.

Table 12

Average Validity Coefficients of the Operationally-derived Variables for IVs for the A/T Criterion

Predictor Variable	Mean Wtd. r	X^2	N
AFQT	.14**	1.64	3875
GCT	.26**	.48	3293[a]
ARI	.21**	5.76	3293[a]
MECH	.04	6.42*	3868[b]
CLER	.14**	4.52	3863
SONR	.08*	3.40	3790
RADO	.12**	6.51	3482
ETST	.16**	9.10*	3248[a,b]
SHOP	.10**	14.94**	3864[b]
YRBI	-.01	24.65**	3310[b]
HS	.07*	4.17	3441[a]
BL	.08*	5.68	3274[a]
YRED	.10**	20.35**	2961[b]
H4	.11**	1.35	3441[a]

[a]Weighted mean based on data from only three samples.

[b]Because of the large number of nonhomogeneous rs, it was decided to include them in the analysis. These means are based on nonhomogeneous rs.

*$p < .05$.

**$p < .01$.

Table 13 presents the experimental variables that had statistically significant validities for the A/T code. As shown, 37 experimental variables significantly predicted the rating progression of IVs (i.e., they predicted whether they would advance into technical ratings). In general, the predictive accuracies of experimental tests were considerably greater for the A/T than for the job performance criterion. The highest significant coefficients were for (1) the Military, Mathematics, and Achievement Interest scales of the SVIB, (2) the Memory for Numbers Tests (total correct), and (3) the AFQT Arithmetic Reasoning subtest score.

Shrunken multiple regression coefficients for the three predictor sets used with the four phase samples are shown in Table 14. The shrunken Rs for the operational selection battery ranged from .11 to .33. Education and racial group variables and AFQT incremented the shrunken R from the operational tests by one to eight points. Only for Phase 3 did experimental tests serve to increase the shrunken R available from the complete set of operational variables.

GCT and CLER were the operational tests that were the most effective for predicting rating progression for IVs. These two tests were present in the sets selected for three of the four phases. Experimental variables that served to increment the shrunken Rs available from operational variables were the Mathematics, Military Interest, Achievement, and Science scales of the SVIB. These findings suggest that the personal characteristics that are the most important determinants of the promotability of Mental Group IVs are verbal and perceptual speed abilities and appropriate personal interests.

Although variation occurred from sample to sample, in general these data suggest that operational tests predict rating progression of IVs with a validity coefficient of about .30, and biographical variables add about five to seven points to the predictiveness of operational tests. Based on the Phase 1 and Phase 3 findings, scales from the SVIB and Manual Speed Tests may be useful for incrementing the operational valdities to bring the total maximum validity to .40 or .45. Thus, the maximum accuracy of prediction of rating progression of IVs ranges from 5 to 15 points higher than the maximum accuracy achievable from prediction of on-job performance.

Comparison of the Length of Service and Attrition Characteristics of Hi- and Lo-IVs

Table 15 shows that the length of service and attrition characteristics of Hi- and Lo-IVs were very similar. Although some of the differences between the two groups were statistically significant, they were not of practical significance.

On the average, Hi-IVs remained in the Navy about 2 months (10%) longer than did Lo-IVs. Also, 18.6 percent of Hi-IVs, versus 14.1 percent of Lo-IVs, remained in the Navy longer than 2 years. After 4 years, however, the retention rates for the two groups were essentially the same. Similarly, there were only small differences between the groups in percentage of attrition discharges.

Table 13

Experimental Variables Having Significant Zero-order Validities for IVs for the A/T Criterion

Experimental Variable	r	N
Phase 1		
Memory for Numbers Test		
Total Correct	.25**	358
Weighted Total Correct	.22**	442
AFQT Retest		
Vocab./Verbal Subtest Score	.17**	449
Arith. Reas. Subtest Score	.22**	449
Total Score	.19**	449
Hand Skills Test		
Part 3 Minus Base Rate	.17**	431
Total Score	.20**	432
Manual Speed Test		
Part 3 Minus Base Rate	.16**	416
Total Score	.14**	415
Phase 2		
Dominoes Test Score	.20**	382
Maze Test Score	.17**	252
Listening Skills Test Score	.17**	283
BIF Items		
Family financial position in comparison with classmates	-.10**	562
Number of siblings who went to college	.10**	514
Job Check List Scales		
Communications	.08*	722
Verbal Activities	.09*	722
Clerical Hand Skills	.11**	722
Phase 3		
SVIB Scales		
Business Management	.15**	418
Merchandising	.16**	418
Office Practices	.19**	418
Military	.26**	418
Mathematics	.30**	418
Science	.16**	418
Teaching	.15**	418
Achievement	.28**	418
Social Introversion	-.18**	418
Phase 4		
Coding Test Score	.08**	1889
Card Pattern Test Score	.07*	1145
Word Finding Test Score	.11**	1890
BIQ Scales		
Logically Derived:		
Home Activities and Responses	.14**	1011
Recreation, Sports, and Hobbies	.09**	1011
Economic and Financ. Respon.	.14**	1011
Interpersonal Interaction	.08**	1011
Socioeconomic Status	.09**	1011
Perceived Social Status	.08*	817
Factorially Derived:		
Activities and Experiences	.11**	1011
Social Leadership	.10**	775

*$p < .05$.
**$p < .01$.

Table 14

Predictors Providing Significant Validity Increments to the Shrunken, Stepwise Multiple Rs of IVs for the A/T Criterion

Set of Predictors	Phase 1			Phase 2			Phase 3			Phase 4		
	Shrunken R	Variable Selected	N	Shrunken R	Variable Selected	N	Shrunken R	Variable Selected	N	Shrunken R	Variable Selected	N
1. Operational Navy Classification Tests	.23	GCT	445	.26	ARI	917	.11	RADO	456	.26	GCT	1876
	.29	CLER	442	.29	ETST	917				.27	CLER	1876
	.32	SHOP	441	.31	GCT	917						
				.33	CLER	917						
2. Variables in set (1) plus AFQT, YRBI, YRED, HS, and BL	.23	YRED	449	.26	ARI	917	.13	AFQT	579	.26	GCT	1876
	.30	ARI	445	.29	ETST	917	.16	RADO	456	.27	BL	1876
	.33	AFQT	435	.31	GCT	917	.18	SHOP	456	.28	YRBI	1876
	.35	CLER	435	.32	YRBI	917						
	.37	SHOP	435	.33	BL	917						
	.40	YRBI	435									
3. Variables in set (2) plus the experimental tests	.23	YRED	449	.26	ARI	917	.28	SVIB Math Scale	418	.26	GCT	1876
	.30	ARI	445	.29	ETST	917				.27	BL	1876
	.33	Manual Spd. Test Diff. Score	416	.31	GCT	917				.28	YRBI	1876
				.32	YRBI	917	.31	SVIB Mil. Int. Scale	418			
				.33	BL	917						
	.36	SHOP	416				.32	AFQT	418			
	.38	Digit Span Total Correct	358				.34	SVIB Achievement Scale	418			
							.36	RADO	418			
							.37	SVIB Science Scale	418			

Table 15

Length of Service and Attrition Statistics
for Hi- and Lo-IVs in the Total Sample

Criterion	Lo-IV (N=2476)	Hi-IV (N=1967)
1. Average Length of Service (mos.)	21.9	23.9***
2. Percent discharged with less than:		
12 mos. service	6.5	5.2
14 " "	11.9	9.1*
16 " "	45.0	40.6**
18 " "	68.5	63.0***
20 " "	84.1	79.1***
24 " "	85.9	81.4***
48 " "	93.9	93.1
130 " "	99.1	99.1
3. Percent of attrition discharges[a]	12.7	11.2

Note. Statistically significant differences between Lo- and Hi-IVs on variables are indicated by asterisks after the Hi-IV column entries.

[a]The categorizations of the Navy Enlisted History Tape were used to define attrition (i.e., any nonnormal termination of enlistment, such as a discharge for misconduct, unsuitability, etc.). Normal discharges are those occurring on expiration of enlistment, release to fleet reserve, etc.

*$p < .05$.
**$p < .01$.
***$p < .001$.

Racial Group

Predictors of On-job Performance of Blacks and Non-Blacks

1. Operational Predictors

The major reasons for comparing Blacks and non-Blacks were to determine whether or not (a) the validity coefficients of operational tests were statistically significant for both racial groups and (b) the coefficients were significantly higher for one or the other racial group.

Table 16 shows that operational variables are not as useful for predicting on-job performance for Blacks as they are for non-Blacks. In the apprenticeship group, only four of the 27 validity coefficients (15%) of the operational tests were statistically significant for Blacks and no variable had coefficients that were significant in more than one sample. In contrast, all 27 validity coefficients of the operational tests were statistically significant for non-Blacks. Three of the eight coefficients for biographical variables (37.5%) were statistically significant for Blacks compared to seven (82.5%) for non-Blacks.

Similar differences in the incidence of significant validities of operational variables were also characteristic of Black and non-Black samples for the technical group. Only five of the 27 coefficients for operational tests (18%) and one of the nine coefficients for biographical variables (11%) were statistically significant for Blacks in the technical group.

It was possible that the relative lack of significant validity coefficients for Blacks may have been artifactual to some extent because the coefficients for Blacks were frequently larger (for some variables as much as twice as large) than those for non-Blacks. The large sizes of the non-Black samples indicated that the predictive relationships were reliable but rather low. However, because the Black coefficients were based on small samples, the possibility of their chance causality could not be ruled out.

To provide more definitive estimates of the validities of the operational variables for Blacks, $\underline{X}^2$ tests of homogeneity of the coefficients were made and average coefficients were computed. The resulting statistics are shown in Table 17. The last column of the table provides a categorization of the level of statistical significance of the differences between the average $\underline{rs}$ of Blacks and non-Blacks. Superscripts identify means that are based on fewer than three samples or on nonhomogeneous $\underline{rs}$. The means associated with these latter coefficients should be viewed with caution because they are averages of $\underline{rs}$ that were statistically distinct.

The average coefficients for two of the nine operational tests (22%) and all three of the biographical variables (100%) were statistically significant for Blacks in the apprenticeship ratings. Three of the operational tests (33%) and none of the biographical variables had significant validities for Blacks in technical ratings. RADO and MECH were the only operational tests having significant validity coefficients for apprenticeship Blacks. The tests which were valid for Blacks in technical ratings--MECH, ETST, and SHOP--cover specialized knowledges and reasoning abilities.

Table 16

Zero-order Validities of Operationally-derived Variables for On-job Performance of Blacks and non-Blacks

Predictor Variable	Phase 2				Phase 3				Phase 4			
	Bl		N-Bl		Bl		N-Bl		Bl		N-Bl	
	r	N	r	N	r	N	r	N	r	N	r	N
Apprenticeship Group												
AFQT	.05	110	.16**	1290	.12	123	.23**	970	.00	168	.18**	1044
GCT	.01	110	.15**	1209	.24**	124	.20**	982	.02	169	.16**	1033
ARI	.06	110	.18**	1209	.35**	124	.22**	982	.04	169	.17**	1033
MECH	.19*	110	.14**	1209	.12	124	.19**	982	.06	169	.16**	1033
CLER	.09	110	.13**	1209	.06	124	.18**	981	.12	169	.14**	1033
SONR	-.03	110	.06*	1209	.04	123	.13**	973	.01	169	.10**	1033
RADO	.11	110	.12**	1209	.17	100	.12**	912	.23**	149	.14**	966
ETST	.03	110	.11**	1209	.11	123	.14**	972	-.03	169	.10**	1033
SHOP	.14	110	.14**	1209	.10	124	.16**	974	.01	169	.16**	1033
Year of birth (YRBI)	-.04	110	-.01	1209	--	---	--	---	-.24**	169	-.13**	1044
High School Graduate (HS)	.13	110	.14**	1209	.04	125	.16**	987	.21**	169	.10**	1044
Years of Education (YRED)	.18	37	.12**	837	.08	125	.21**	987	.20**	169	.12**	1044
Technical Group												
AFQT	.11	66	.04	2663	.04	45	.11**	1785	.13	54	.12**	1673
GCT	.19	66	.07*	2663	-.14	45	.07*	1818	.06	54	.14**	1657
ARI	.20	66	.07*	2663	.12	45	.12**	1818	.08	54	.15**	1657
MECH	.21	66	.05	2663	.15	45	.09**	1818	.27*	54	.10**	1657
CLER	.05	66	.09**	2663	-.34*	45	.09**	1817	.04	54	.14**	1657
SONR	.22	66	.02	2663	-.13	45	.06*	1864	-.02	54	-.01	1656
RADO	.12	66	.03	2663	-.20	42	.10**	1663	-.30*	52	.06*	1637
ETST	.16	66	.07*	2663	.29*	45	.16**	1684	.03	54	.11**	1656
SHOP	.17	66	.05	2663	.22	45	.09**	1817	.40**	54	.11**	1657
Year of birth (YRBI)	.08	66	-.09**	2660	--	--	--	----	-.11	54	-.16**	1673
High School Graduate (HS)	.08	66	.06*	2662	-.08	46	.10**	1828	.28*	54	.20**	1673
Years of Education (YRED)	-.15	56	.13**	2589	.15	46	.16**	1628	.09	54	.20**	1673

*$p < .05$.
**$p < .01$.

Note. Because the Phase I sample included very few Blacks, the analyses were based on Phase 2, 3, and 4 samples.

Table 17

Average Validity Coefficients of the Operationally-derived Variables for On-job Performance of Blacks and Non-Blacks

Predictor Variable	Blacks			Non-Blacks			r_2-r_1
	Mean Wtd.[a] $\underline{r}$ (r_1)	$\underline{X}^2$	$\underline{N}$	Mean Wtd. $\underline{r}$ (r_2)	$\underline{X}^2$	$\underline{N}$	
Apprenticeship Group							
AFQT	.05	1.01	401	.19**	3.02	3304	.14**
GCT	.09	4.41	403	.17**	1.58	3224	.08
ARI	.05	.03	279[b]	.19**	1.54	3224	.14*
MECH	.12**	1.14	403	.16**	1.41	3224	.04
CLER	.09	.26	403	.15**	1.52	3223	.06
SONR	.01	.28	402	.09**	2.76	3215	.08
RADO	.18**	.96	359	.13**	.27	3087	-.05
ETST	.03	1.36	402	.12**	.91	3214	.09
SHOP	.08	.72	403	.15**	.30	3216	.07
YRBI	-.16**	2.73	279[b]	-.07*	8.23**	2253[b,c]	.09
HS	.14**	2.11	404	.13**	1.98	3240	-.01
YRED	.15**	1.09	331	.15**	5.46	2868	.00
Technical Group							
AFQT	.10	.21	165	.12**	.03	3458[b]	.02
GCT	.06	2.79	165	.07*	.01	4481[b]	.01
ARI	.14	.45	165	.14**	.70	3475[b]	.00
MECH	.22**	.37	165	.03*	3.11	6138	-.14
CLER	-.03	5.39	165	.10**	3.14	6137	.13
SONR	.07	3.06	165	.02	4.13	6003	-.05
RADO	-.10	5.65	160	.06	5.02	5963	.16*
ETST	.16*	1.67	165	.11**	8.54*	6003[c]	-.05
SHOP	.27**	1.90	165	.08*	4.04	6137	-.19*
YRBI	.00	1.02	120	-.12**	5.17*	4333[c]	-.12
HS	.10	3.22	166	.08*	.18	4490[b]	-.02
YRED	.02	2.54	156	.16**	5.27	6090	.14

Note. For the apprenticeship group, AFQT and ARI r_2-r_1, when corrected for restriction of range on the Black scores, were .11 and .12 respectively, both significant at $p < .05$.

[a]Means were computed from sample $\underline{r}$s weighted by the sample $\underline{N}$s.

[b]Mean was based on data from only two samples.

[c]Mean was computed from nonhomogeneous $\underline{r}$s.

*$\underline{p} < .05$.

**$\underline{p} < .01$.

The statistically significant mean rs for non-Blacks in apprenticeship ratings largely reflect the relationships for these personnel presented in Table 16 and previously discussed. Thus, although there were fewer statistically significant mean validity coefficients (Table 17) for tests in the operational battery for Blacks than for non-Blacks, the percentage of deficits was smaller for the averages than for the original coefficients.

More of the operational variables were statistically significant for non-Blacks than for Blacks, and the relative magnitudes of the coefficients for non-Blacks, in general, tended to be somewhat larger than those for Blacks. For the apprenticeship group, non-Blacks had higher mean validities for eight of the operational variables, compared with three for Blacks. For personnel in technical ratings, the relative superiority of the variables was about evenly split: Six variables were more predictive for Blacks and five for non-Blacks.

Because of the substantial differences between the sizes of the samples of Blacks and non-Blacks, the most appropriate indication of whether the tests are differentially predictive for personnel in the two racial groups is the comparison of the statistical significance of the difference between the validity coefficients for Blacks and non-Blacks. For both rating groups, statistically significant differences occurred in the valdity coefficients for Blacks and non-Blacks on two of the operational tests--AFQT and ARI for apprenticeship personnel and RADO and SHOP for technical personnel. Only in the case of RADO was the validity higher for Blacks. To further evaluate these differences, the standard deviations for Blacks and non-Blacks for the four tests were inspected. For the personnel in the technical group, the standard deviations of RADO and SHOP were approximately the same for Blacks and non-Blacks; for the apprenticeship group, however, the standard deviations of AFQT and ARI for Blacks were only .64 and .74 respectively of those of non-Blacks. To compensate for the differences, the observed validity coefficients of Blacks for AFQT and ARI were corrected for restriction in range and the differences in the validity coefficients were again tested for statistical significance. After adjustment, the validity coefficient for AFQT was still significantly higher for non-Blacks than Blacks--but at the .05 level of significance rather than the .01 level. There was no change in the level of significance for ARI.

Thus, in apprenticeship ratings, validity coefficients for Blacks were consistently smaller and less significant than those for non-Blacks. In the technical ratings, the differnces between Blacks and non-Blacks tended to balance out.

The data in Table 16 indicate that biographical variables, which are present in the operational records, hold promise as predictors of on-job performance of both Blacks and non-Blacks assigned to apprenticeship ratings. However, the two biographical variables measuring education should not be used operationally for selection unless they are confirmed on an independent sample.

The substantial differences in the percentage of significant validities in comparison to the original validity coefficient mean for Blacks points up the danger of drawing conclusions based on small sample sizes. This is a particular problem for studies of test bias for racial minorities because of the relatively small proportion of the total population represented by the minorities. Normal sampling procedures, therefore, are likely to produce samples of minority personnel that are too small to produce reliable statistics.

2. Experimental Predictors

As shown in Table 18, five experimental variables had significant validities for Blacks in the apprenticeship group; and two, for Blacks in the technical group. The seven significant validities represented about five percent of the total number of validities of experimental variables, which is the percentage of significant validities that would be expected from chance alone. In contrast, the validities of all but one of the cognitive experimental tests, and considerable percentages of the vocational interest scales, biographical information variables, and personality scales were statistically significant for non-Blacks in both the apprenticeship and the technical groups.

Most of the negative coefficients in Table 18 are for interest measures or biographical variables. The negative values of variables of this type indicate a direction of interest or orientation rather than the association of lower values of a quantitative predictor with higher values of a criterion. Four of the five coefficients that were statistically significant for apprenticeship Blacks were for interest measures. In contrast, both of the variables with significant validities for Blacks in technical ratings were for cognitive tests and had positive signs (indicating that higher ability on the culture-fair measure was associated with higher on-job performance). From Table 18, it appears that high performing Blacks in the apprenticeship ratings (1) do not believe that something stops them when they try to advance, (2) score high on retention characteristics for the Navy, (3) have low clerical interests, and (4) are extraverts. Distinguishing characteristics of the better performing Blacks in technical ratings were their relatively high scores on Dominoes and Listening Skills 2.

Because of the small number of Blacks in the sample for the technical group, multiple regression comparisons were not made of operational and experimental variables as predictors of on-job performance. Since the number of experimental variables with statistically significant validities for Blacks was not greater than would be expected from chance alone, however, it seems unlikely that adding any of the experimental variables would have increased the shrunken multiple regression coefficients for Blacks.

Predictors for Rating Progression of Blacks and Non-Blacks

The high predictiveness of the operational variables for rating progression of both Blacks and non-Blacks is shown in Table 19. For both racial groups, all 35 coefficients in the table are statistically significant at the .01 level. Cognitive tests in general and AFQT, GCT, and ARI in particular were the best predictors of the A/T criterion.

Weighted averages of the validity coefficients in Table 19 are shown in Table 20. Seventeen percent of the averages for Blacks and 42 percent of the averages for non-Blacks are based on nonhomogeneous rs. The degree to which the respective coefficients of Blacks and non-Blacks correspond, however, both in rank and in magnitude, is striking. For both Blacks and non-Blacks, AFQT, GCT, and ARI had zero-order validities that ranged from the low .50s to the low .60s. Biographical variables were not as effective as the operational tests as predictors for either Blacks or non-Blacks.

Table 18

Significant Zero-order Validities of Experimental Variables[a] for On-job Performance of Blacks and Non-Blacks

	Validity for On-job Performance							
	Apprenticeship Group				Technical Group			
	Bl		N-Bl		Bl		N-Bl	
Experimental Variables	r	N	r	N	r	N	r	N
Phase 2								
Dominoes Test Score	-.20	50	.18**	541	.48**	39	.06*	1171
Listening Skills Test I Score	-.20	32	.13**	391	--	--	--	----
BIF Items								
Father's level or occupation	-.24	48	.12**	695	--	--	--	----
Father's level of education	--	--	--	---	-.27	32	.11**	1778
Both parents in home	.01	69	.14**	760	.14	42	.06*	1856
Number of siblings who went to college	.12	61	.08*	715	--	--	--	----
Level of educational attainment expected of individual by his parents	-.22	66	.09*	687	.00	42	.10**	1771
Opinion of individual that something stops him when he tries to advance	-.33*	53	.03	717	--	--	--	----
Opinion of individual that he has little chance of success in life	.02	48	.08*	697	.02	39	.06*	1813
Special BIF retention scale	.22*	110	.11**	1209	.15	66	.08**	2663
Job Check List Scales								
Clerical	-.22*	137	-.07*	1044	--	--	--	----
Domestic Activities	--	--	--	---	-.15	55	.06*	2442
Clerical Hand Skills	-.22*	137	-.05	1044	--	--	--	----
Phase 3								
Listening Skills Test II Score	.17	41	.22**	435	.44*	21	.07	908
Performance Index								
Personal Maturity Score	.25	16	.18*	185	--	--	--	----
RAQ Scales								
Fringe Benefits and Retirement	--	--	--	---	.74	5	.19**	293
Pay	--	--	--	---	.22	5	-.15**	29[illegible]
Advancement	-.24	18	-.19**	210	--	--	--	-- --
SVIB Scales								
Business Management	--	--	--	---	.07	42	.07*	1780
Merchandising	--	--	--	---	.06	42	.08**	1780
Office Practices	--	--	--	---	-.03	42	.08**	1780
Military	-.06	87	.07*	885	--	--	--	----
Technical Supervision	--	--	--	---	.09	42	.10**	1780
Mathematics	.02	87	.09**	885	.23	42	.09**	1780
Mechanical	--	--	--	---	.09	42	.07*	1780
Navy Interest Scale	.05	87	.08*	885	-.22	42	.09**	1780
Liberal, Conservative	--	--	--	---	-.15	42	.10**	1780
Masculinity/Femininity	--	--	--	---	.16	42	.09**	1780
Social Introversion	-.23*	87	-.01	885	--	--	--	----
Phase 4								
Coding Test Score	.12	169	.15**	1043	.11	54	.12**	1673
Word Finding Score	.08	169	.18**	1044	-.02	54	.11**	1673
Mechanical Principles Test Score	-.12	73	.17**	505	.43	14	.08*	750
BIQ Scales								
Logically Derived:								
Interpersonal Interaction	--	--	--	---	-.12	21	.07*	974
Factorially Derived:								
Social Leadership	--	--	--	---	-.33	20	.09**	921
Recruit Temperament Survey (RTS)	-.22	68	-.13**	394	-.16	29	-.08*	671

[a]Data are shown for both Blacks and non-Blacks for any cell for which either racial group had significant validities. Data for experimental variables with no significant validities are not shown.

*p < .05.

**p < .01.

Table 19

Zero-order Validities of Operationally-derived Variables for the A/T Criteria for Blacks and Non-Blacks

Predictor Variable	Phase 2				Phase 3				Phase 4			
	Black		Non-Black		Black		Non-Black		Black		Non-Black	
	r	N	r	N	r	N	r	N	R	N	R	N
AFQT	.60**	290	.57**	5019	.58**	170	.45**	2783	.58**	452	.64**	3886
GCT	.59**	290	.56**	5019	.54**	171	.39**	2829	.49**	451	.63**	3655
ARI	.58**	290	.52**	5019	.54**	171	.37**	2829	.49**	451	.58**	3655
MECH	.43**	290	.34**	5019	.53**	171	.24**	2829	.25**	451	.41**	3655
CLER	.25**	290	.23**	5019	.32**	171	.17**	2827	.19**	451	.24**	3655
SONR	.35**	290	.22**	5019	.34**	170	.19**	2685	.22**	450	.31**	3654
RADO	.28**	290	.28**	5019	.40**	144	.27**	2601	.25**	390	.35**	3487
ETST	.49**	290	.50**	5019	.46**	170	.43**	2684	.36**	450	.53**	3654
SHOP	.41**	290	.38**	5019	.43**	171	.27**	2820	.30**	451	.46**	3655
YRBI	-.14*	289	-.10**	5015	--	---	--	----	-.12**	453	-.18**	3686
HS	.28**	290	.33**	5019	.20**	173	.24**	2844	.13**	453	.33**	3686
YRED	.28**	137	.25**	4124	.27*	173	.25**	2844	.20**	453	.38**	3686

*$p < .05$.
**$p < .01$.

Table 20

Average Validity Coefficients of the Operationally-derived Variables for the A/T Criterion

Predictor Variable	Blacks			Non-Blacks		
	Mean Wtd. r	X^2	N	Mean Wtd. r	X^2	N
AFQT	.58**	.15	912	.60**	28.96**	8705[a,b]
GCT	.53**	3.48	912	.60**	1.01	8674[a]
ARI	.53**	2.77	912	.54**	15.09**	8674[a,b]
MECH	.37**	16.23**	912[b]	.38**	13.78**	8674[a,b]
CLER	.24**	2.46	912	.24**	.21	8674[a]
SONR	.28**	4.19	910	.21**	1.24	7704[a]
RADO	.29**	2.95	824	.28**	.21	7620[a]
ETST	.42**	4.83	910	.52**	3.55	8673[a]
SHOP	.41**	7.01**	461[a,b]	.38**	76.86**	11494[b]
YRBI	-.13**	.03	742	-.13**	14.19**	8701[b]
HS	.19**	4.29	916	.33**	.00	8705[a]
YRED	.24**	2.25	763	.25**	.00	6968[a]

[a]Wtd. mean based on data from only two samples.

[b]Mean was computed from nonhomogeneous rs.

*p < .05.
**p < .01.

The experimental variables that had significant validities for the rating progression of either Blacks or non-Blacks are shown in Table 21. Nearly two and a half times as many of the variables were predictive for non-Blacks as for Blacks (56 vs. 23). In general, as was true for the operational variables, the better experimental predictors tended to be cognitive tests. Noncognitive variables, in addition to having lower validity coefficients than cognitive variables, tended not to be predictive for one or the other of the racial groups. For experimental variables with relatively large sample sizes, the best predictors of rating progression, for both Black and non-Blacks, were (1) scores of the Dominoes, Listening Skills I and II, and Word Finding Tests and (2) three SVIB scales--Achievement, Science, and Mathematics.

Multiple regression comparisons of the operational and experimental predictor sets, shown in Table 22, illustrate still another aspect of similarity in the predictabilities of rating progression of Blacks and non-Blacks and in the variables that were included in the subsets of maximally predictive variables. For every regression, regardless of racial group, AFQT and GCT and ARI of the Navy classification tests were the best predictors. Despite the fact that a large number of experimental variables had statistically significant zero-order validity coefficients for rating progression of both Blacks and non-Blacks, only one experimental variable added to the predictiveness of the operational variables for each racial group. In each case, the variable provided about a two percent increase to the predictiveness of the composite.

In general, the findings indicate that an optimally weighted subset of Navy classification tests predicts rating progression of Blacks and non-Blacks with validities of about .58 to .60. The addition of AFQT would increase the predictiveness of the subset two or three points on the average, making the maximum predictability of rating progression about .61 or .63. An experimental interest measure may add one point to the optimally weighted subset of operational variables.

The types of variables that were most predictive of rating progression of the full range samples of Blacks and non-Blacks contrast somewhat with those which were most predictive of rating progression of Mental Group IVs. For the full range samples, the best predictors were AFQT, GCT, and ARI, tests of generalized mental ability. In contrast, for IVs, the best predictors included both cognitive tests and interest measures. Furthermore, the most predictive cognitive tests for IVs measured applied abilities as well as generalized mental abilities.

Table 21

Significant Zero-order Validities of Experimental Variables for Blacks and Non-Blacks for the A/T Criterion

Experimental Variables	Black		Non-Black	
	r	N	r	N
Phase 2				
Dominoes Test Score	.54**	143	.39**	2210
Maze Test Score	.19	68	.27**	1241
Listening Skills Test I Score	.40**	79	.43**	1568
BIF Items				
Father's level of occupation	.05	132	.08**	3180
Father's level of education	.03	131	.06*	3207
Did parents live together during most of school years?	-.01	183	.09**	3403
During the school years, the relative affluence of the family in relation to those of classmates	-.15*	181	-.09**	3401
Number of brothers and sisters who have gone to college	-.11	159	.10**	3211
Number of brothers and sisters who dropped out of high school before finishing	-.01	156	18**	3215
Educational attainment expected of the individual by his parents or guardians	-.05	176	.23**	3298
Number of books read on own time during the last three months	.07	177	.12**	3389
Good luck is more important than hard work for success	.03	158	.13**	3315
Every time I try to get ahead, something or somebody stops me	.17*	155	.16**	3306
People like me don't have much of a chance to be successful in life	.13	148	.18**	3249
Special BIF Retention Scale	.11	290	.18**	5019
Socioeconomic Status	05	290	.16**	5019
Job Check List Scales				
Clerical	.24**	219	.04	4528
Domestic Activities	.17*	219	.15**	4528
Communications	.07	219	-.12**	4528
Verbal Activities	.13	219	-.08	4528
Graphics	-.03	219	-.09**	4528
Phase 3				
Listening Skills II Score	.33**	62	.24**	1352
Performance Index				
Personal Maturity Score	.24	29	.27**	694
General Maturity Score	.43*	29	.24**	694
RAQ Scales				
Equipment and Facilities	.31	24	.10*	510
Patriotism and Duty	.42*	24	-.02	510
SVIB Scales				
Business Management	.04	130	.08*	2691
Merchandising	-.04	130	.08*	2691
Office Practices	.01	130	.07*	2691
Military	.03	130	.06*	2691
Mathematics	.30**	130	.27**	2691
Science	.28**	130	.22**	2690
Mechanical	.10	130	.08*	2691
Nature	-.01	130	.07*	2691
Teaching	.02	130	.14**	2691
Navy Interest	.18*	130	.15**	2691
Achievement	.36**	130	.26**	2691
Social Introversion	-.15	130	-.15**	2691
Occupational Level	.20*	130	.18**	2691

*p < .05.
**p < .01.

Table 21 (Continued)

Experimental Variables	Black		Non-Black	
	r	N	r	N
Phase 4				
Coding Test Score	.16**	453	.24**	3685
Card Pattern Test Score	.23**	275	.41**	2014
Word Finding Test Score	.27**	453	.27**	3686
Mechanical Principles Test Score	.21**	178	.46**	1672
BIQ Scales				
Logically Derived:				
Impoverishment of Home Environment	.00	244	-.21**	2150
Teenage Activities	.16**	244	.23**	2150
Home Activities and Responsibilities	.16**	244	.17**	2150
Mechanical and Technical Experiences	.11	244	.18**	2150
Recreation, Sports, and Hobbies	.11	244	.16**	2150
Literacy/Cultural Interests	.11	244	.19**	2150
Travel Experiences	.07	244	.07*	2150
Economic and Financial Responsibilities	.11	244	.18**	2150
Interpersonal Interaction	.21**	244	.22**	2150
Socioeconomic Status	.09	244	.16**	2150
Perceived Social Class	.10	244	.20**	2150
Factorially Derived:				
Activities and Experiences	.13	163	.08*	1889
Electro/Mechanical	.18	163	.11**	1889
Social Leadership	.18*	163	.10**	1889
Recruit Temperment Survey (RTS)	-.09	190	-.22**	1448

*$p < .05$.
**$p < .01$.

Table 22

Variables Providing Significant Validity Increments to the Shrunken Multiple Rs of Blacks and Non-Blacks for the A/T Criterion

Set of Predictors	Blacks			Non-Blacks		
	Shrunken R	Variable Selected	N	Shrunken R	Variable Selected	N
Phase 2						
1. Operational Navy Classification Tests	.58	GCT	290	.55	GCT	5019
	.62	ARI	290	.57	ARI	5019
	.64	MECH	290	.58	MECH	5019
2. Variables in Set (1) plus AFQT, YRBI, YRED, and HS	.59	AFQT	290	.57	AFQT	5019
	.62	GCT	290	.60	GCT	5019
	.64	YRED	137			
	.65	MECH	137			
3. Variables in Set (2) plus the Experimental Tests	.59	AFQT	290	.57	AFQT	5019
	.62	GCT	290	.60	GCT	5019
	.64	YRED	137			
	.65	MECH	137			
Phase 3						
1. Operational Navy Classification Tests	.53	ARI	171	.43	ETST	2684
	.59	MECH	171	.48	RADO	2684
2. Variables in Set (1) plus AFQT YRBI, YRED, and HS	.57	AFQT	170	.45	AFQT	2783
	.61	MECH	170	.48	ETST	2684
	.62	ARI	170			
3. Variables in Set (2) plus the Experimental Tests	.57	AFQT	170	.45	AFQT	2783
	.61	MECH	170	.48	ETST	2684
	.62	ARI	170	.49	OF[a]	2684
Phase 4						
1. Operational Navy Classification Tests	.47	GCT	451	.62	GCT	3655
	.54	ARI	451	.64	ARI	3655
2. Variables in Set (1) plus AFQT, YRBI, YRED, and HS	.57	AFQT	452	.64	AFQT	3686
	.59	ARI	451	.66	GCT	3655
	.60	GCT	451			
3. Variables in Set (2) plus the Experimental Tests	.57	AFQT	452	.64	AFQT	3686
	.59	ARI	451	.66	GCT	3655
	.60	WF[b]	451			

[a] SVIB Scale OF.

[b] Word Finding.

DISCUSSION AND CONCLUSIONS

In general, analyses utilizing shrunken multiple regression coefficients for estimating the potential increases that experimental variables would add to the validity of a battery of operational selection variables tend to overestimate the increase. This occurs because the effects of chance correlations on the selection of variables by the stepwise process are not fully compensated for by shrinkage formulae. Therefore, application of the regression equations to a new sample (i.e., cross validation) is preferred for estimating the impact of the addition of predictor variables on the validity of a selection battery.

Shrunken multiple regression estimates of validity increments were nevertheless used for the present study because the sample sizes were too small to permit cross validation. These estimates were derived using a very conservative correction formula that has been shown in other research to accurately approximate the actual cross validated coefficients.

Summaries of the findings and conclusions are presented separately for the mental level and the racial group categorizations of the samples.

Mental Level Group

1. Mental Group IVs serving as enlisted personnel in the Navy were primarily assigned to apprenticeship ratings. Only about 18 percent of the IVs succeeded in attaining and remaining in technical ratings during their first enlistment. Since the study encompassed a period in which a maximum effort was being carried out to place IVs in technical ratings, it appears that IVs lack aptitudes for most technical ratings, or at least lack aptitudes in comparison with those of the I-IIIs, with whom they must compete.

About 29 percent of the IVs in technical ratings were Stewards. Other ratings in the technical group which had the largest percentages of IVs were Equipment Operator, Commissaryman, Hull Maintenance Technician, Yeoman, Engineman, Boatswain's Mate, Hospital Corpsman, Builder, Radioman, and Machinist's Mate. IVs in those 11 ratings constituted 72 percent of the IVs in the technical group.

If the input of IVs were limited to those with higher mental ability, and a 4-year enlistment period were used, the percentages of IVs advancing into technical ratings would be expected to increase.

2. Twenty-one percent of the IVs in the technical group were in ratings having shortages in retention of career personnel. Since other research has shown that larger proportions of Mental Group IVs than of I-IIIs are retained in some technical ratings, IVs may have good retention rates in some of the shortage ratings.

In a study of the retention rate for enlisted personnel classified by mental group, Cory (1971) found higher than average retention rates for IVs in the Hull Technician, Machinist's Mate, Boiler Technician, Ship's Serviceman, Aviation Machinist's Mate, Aviation Structural Mechanic, and Gunner's Mate ratings. This suggests that a closer look should be given to the IVs who have been appointed to these ratings over the years to determine just what characteristics of Mental Group IVs are required for success. Such research would also permit a closer analysis of any tradeoffs between on-job performance and retention capabilities that might be associated with appointment of larger numbers of IVs to these ratings.

3. There are no substantial differences in the length of service characteristics of Hi- and Lo-IVs. The mean length of service for personnel in each group was slightly more than 20 months, and less than 20 percent of each group had more than 2 years of enlisted service. The latter finding largely reflects the fact that most of these personnel had 2-year enlistments.

4. The average job performance of Mental Group IVs was consistently rated lower than that of I-IIIs by their immediate supervisors. In general, the average marks of IVs ranged from the 33rd to the 39th percentiles of the distribution of job performance marks (the average performance being the 50th percentile).

5. Most of the operational tests were significant predictors of the on-job performance of both IVs and I-IIIs in apprenticeship and technical groups. The Radio Code Aptitude, Mechanical, and Clerical tests were the best predictors for on-job performance of IVs in the apprenticeship group; and the Armed Forces Qualification, Electronics Selection, and Arithmetic tests, the best predictors for IVs in the technical group. Variables measuring amount of education were better predictors of on-job performance of IVs in the technical groups than any of the operational tests.

6. A number of experimental tests were significantly predictive of on-job performance of IVs in both the apprenticeship and technical groups. The experimental tests, however, did not add enough to the predictive accuracy already available from operational predictors to justify their use operationally for the selection of Mental Group IV personnel.

7. Generally, for apprenticeship IVs, the maximum predictiveness of an optimally weighted battery of operational classification tests was .15 to .20. Biographical variables added to the predictor set would improve the maximum predictiveness about .03 or .04. The addition of scales for personality, interests and neuroticism might increase the maximum predictiveness of the resulting battery to about .30 for the apprenticeship IVs.

8. The findings suggest that high school graduation plus the scores from one or two of the operational tests will predict the on-job performance of IVs in technical ratings with a shrunken $\underline{R}$ of about .25 and selected experimental variables added to the battery could increase the predictiveness five to ten points. The maximum predictiveness of an optimally-weighted battery of operational and experimental variables for on-job performance of IVs in technical ratings might be expected to be in the .30 to .40 range. However, because of the relatively small sizes of the samples of IVs in technical ratings, these findings should be replicated before they can be considered reliable.

9. Operational variables were more predictive of rating progression than of on-job performance of Mental Group IVs. Tests measuring generalized and applied mental abilities were both among the best predictors of rating progression of IVs.

10. Four scales of the Strong Vocational Interest Battery--Mathematics, Military, Achievement, and Science--were among the best experimental predictors of rating progression of IVs.

11. In general, the maximum predictiveness of an optimal battery of operational tests for the rating progression of IVs appears to be about .30. Appropriate measures of personal and vocational interests from the Strong Vocational Interest Battery could probably increase this maximum to .40 or .45.

12. Despite an extensive and wide-ranging effort to develop and validate tests that were more appropriate for use in selecting and assigning Category IV personnel than the operational classification tests, usable results were very meager. In fact, the so-called "culture fair" tests generally were less valid as predictors of on-job performance of IVs than the verbally loaded tests in the operational battery.

13. Because of the extensive literature search carried out for this project, it seems unlikely that promising paper-and-pencil tests that offer potential improvement in ability to identify capable Mental Group IVs have been overlooked. The most reasonable conclusion is that substantial breakthroughs in the use of paper-and-pencil methodology to identify low scoring personnel with high job performance abilities and potential for rating progression are unlikely. Future development of computerized measurement of abilities, together with branching test construction, might provide a more accurate method of identifying these personnel. Psychobiological measures may also prove useful (Lewis, Rimland, & Callaway, 1976). The classification tests used operationally, however, are likely to remain the best written tests for screening Mental Group IV personnel in the near future.

Racial Group

To some extent the findings and recommendations for Blacks resemble those for IVs; there is a substantial overlap among Black and IV personnel in the apprenticeship and technical groups. However, because the constraints on these groups differ, the impacts of the findings and, consequently, the recommended courses of action for Blacks and IVs also differ.

In the case of Mental Group IVs, the important aspect is the total predictive accuracy that is achievable for Mental Group IVs. Thus, the validities of experimental tests are compared with those of operational tests with an objective of possibly substituting one for the other. The interest is not only on the total magnitude of the predictive validity obtainable, but also on the tradeoff between time and expense of testing and incremental validity.

Although these considerations are also important for comparisons of Blacks and non-Blacks, they are subordinate to two other standards. Comparisons of Blacks and non-Blacks are carried out primarily to determine (1) whether or not the validities of operational tests for criteria are statistically significant and (2) whether or not these validities differ significantly from one racial subgroup to the other.

The first of these standards is stipulated in Title 41 of the Equal Employment Opportunity Act as a requirement for the unbiased use of tests for personnel selection. The second is contained in the Standards for Educational and Psychological Tests (American Psychological Association, 1974).

Thus, for Mental Group IVs, the important consideration is the practical utility of the validity coefficients. For Blacks versus non-Blacks, the statistical significance of the validity coefficients is equally, if not more important.

1. More of the operational variables were found to be predictive of on-job performance of non-Blacks than of Blacks. Job performance of Blacks in apprenticeship ratings was significantly predicted by 22 percent of the operational tests and by both binary high school graduation and years of education. Job performance of Blacks in technical ratings was significantly predicted by 33 percent of the operational tests. In comparison, all of the operational variables were significantly predictive of on-job performance of non-Blacks in apprenticeship ratings and 10 of them (71%) significantly predicted on-job performance of non-Blacks in technical ratings.

2. Although many more of the variables were significantly predictive for non-Blacks than for Blacks, these differences were due largely to the fact that the samples of non-Blacks were much larger than those of Blacks and, therefore, resulted in greater statistical reliability of the findings. The actual magnitude of the validity coefficients was generally not substantially greater for non-Blacks than for Blacks. The difference between non-Blacks and Blacks in predictiveness of the operational variables were somewhat greater for the apprenticeship than for the technical groups.

3. Operational tests that were significantly less predictive of job performance of Blacks in apprenticeship ratings than of non-Blacks in those ratings were AFQT and ARI. For personnel in the technical group, the Radio Code Aptitude Test was significantly more predictive for Blacks; and the Shop Practices Test, for non-Blacks.

4. Similarly, many more of the experimental variables significantly predicted on-job performance of non-Blacks than of Blacks. It is unlikely that any of the experimental variables would significantly improve the predictiveness for on-job performance of Blacks of an optimally-weighted battery of operational variables.

5. For both Blacks and non-Blacks, operational variables were much more predictive of rating progression than of on-job performance. The AFQT, the General Classification Test, and the Arithmetic Test were the most predictive operational variables for rating progression, and the predictive relationships for these tests were very similar for Blacks and non-Blacks. Variables measuring level of education were less effective as predictors of rating progression than of on-job performance.

6. When multiple regression coefficients were computed for rating progression for Blacks and non-Blacks, experimental variables generally failed to increase the predictiveness of the operational tests.

7. In general, an optimally weighted subset of Navy classification tests predicts rating progression of both Blacks and non-Blacks at validities of about .58 to .60. The addition of the AFQT to this subset increases its predictiveness two or three points for a maximum predictability for rating progression of both Blacks and non-Blacks of .61 or .63.

8. Measures of age and education that are available operationally had statistically significant validities for both Black and non-Black personnel in apprenticeship ratings. Furthermore, the sizes of their validities indicate that these variables would aid in the classification of Navy enlisted personnel.

9. The test of validity used for the present study was rigorous. The 20 to 21 month time interval that elapsed between testing and collection of the supervisory marks is considerably longer than is necessary to learn to perform competently in most Navy ratings, and it is also longer than is customary for studies of test bias. It would be expected that this increase in time interval would serve to increase the normal attenuation of validity coefficients that occurs over time.

RECOMMENDATIONS

1. None of the experimental tests included in the present research should be used operationally for selection and/or classification at this time. However, future exploratory development concerned with potential modifications of the operational tests should, wherever feasible, include variables which were found to have significant validities in the present study.

2. Mental Group IVs should continue to be selected for service in the Navy on the basis of their scores on the operational classification battery.

REFERENCES

Abrahams, N. M., & Neumann, I. The validation of the Strong Vocational Interest Blank for predicting Naval Academy disenrollment and aptitude (STB 73-03). San Diego: Naval Personnel and Training Research Laboratory, February 1973.

Abrahams, N. M., Neumann, I., & Githens, W. H. The Strong Vocational Interest Blank in predicting NROTC officer retention: Part I. Validity and reliability (STB 68-7). San Diego: Naval Personnel and Training Research Laboratory, January 1968.

Bowser, S. E. Non-cognitive factors as predictors of individual suitability for service in the U.S. Navy (NPRDC Tech. Rep. 74-13). San Diego: Navy Personnel Research and Development Center, April 1974. (AD-780 438)

Cattell, R. B. Personality and motivation: Structure and measurement. New York: Harcourt and Brace, 1957.

Cory, C. H. A comparison of retention of category IVs and non-IVs in fifty-eight Navy ratings (NPTRL SRR 71-13). San Diego: Naval Personnel and Training Research Laboratory, March 1971.

Cory, C. H. A comparison of the job performance and attitudes of category IVs and I-IIIs in sixteen Navy ratings (NPRDC Tech. Rep. 76-35). San Diego: Navy Personnel Research and Development Center, 1976. (AD-A024 642)

Curtis, E. W. Prediction of enlisted performance. I. Relationship among aptitude tests, Navy school grades, the report of enlisted performance evaluation and advancement examinations (NPTRL TB STB 71-10). San Diego: Naval Personnel and Training Research Laboratory, 1971.

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Edgerton, H. A., & Sylvester, R. W. Construction of a maturity scale (Contract 15-67-C-0188). Washington, DC: Performance Research, Inc. 1967.

Ghiselli, E. E. The validity of occupational aptitude tests. New York: Wiley, 1966.

Gough, H. G., & Domino, G. The D-48 test as a measure of general ability among grade school children. Journal of Consulting Psychology, 1963, 27, 344-349.

Jensen, A. R. Individual differences in learning: Interference factor (U.S. Office of Education Cooperative Research Project No. 1867). Berkeley: University of California, 1964.

Jensen, M. B. The low-level airman in retesting and basic training: A sociopsychological study. Journal of Social Psychology, 1961, 55, 177-190.

Kipnis, D. A noncognitive correlate of performance among lower aptitude men. Journal of Applied Psychology, 1962, 46(1), 76-80.

Lau, A. W., & Abrahams, N. M. A comparison of two interest inventories for use with enlisted personnel (NPTRL SRM 71-12). San Diego: Naval Personnel and Training Research Laboratory, May 1971.

Lewis, G. M., Rimland, B., & Callaway, E. Psychobiological predictors of success in a Navy remedial reading program (NPRDC Tech. Rep. 77-13). San Diego: Navy Personnel Research and Development Center, December 1976. (AD-A037 339)

Naval Health Research Center. Navy Enlisted Service History Documentation (Revised 6/79).

Porteus, S. D. Porteus Maze Test: Fifty years application. Palo Alto: Pacific Books, 1965.

Raines, G. N., Wittson, C. L., Hunt, W. A., & Herman, R. S. Psychiatric selection for military service. Journal of American Medical Association, 1954, 156, 817-821.

Ratliff, F. R. & Earles, J. A. Research on the management, training, and utilization of low-aptitude personnel: An annotated Bibliography (AFHRL TR-76-69). Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, December 1976.

Standards for educational and psychological tests. Prepared by a joint committee of the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. Frederick B. Davis, Chair. Published by the American Psychological Association, Inc., Washington, DC, 1974.

Standlee, L. S., & Saylor, J. C. Feasibility of equipment operator class "A" school training for Group IV personnel (NPRA SRR 68-21). San Diego: Naval Personnel Research Activity, January 1969.

Waite, R. R., & Barnes, E. H. Predicting success or failure in the Navy, A longitudinal study. U. S. Navy Medical Corps Report (Unpublished).

APPENDIX

DESCRIPTIONS OF THE 19 EXPERIMENTAL TESTS

DESCRIPTIONS OF THE 19 EXPERIMENTAL TESTS

Cognitive Skills

The cognitive skills tests described below measured "fluid" rather than "crystallized" abilities; that is, those involving problem-solving capabilities and quickness and response rather than learned or acquired abilities (Cattell, 1957). Further, they generally dealt with concrete problems rather than abstract problems or mental operations. The two tests that did involve abstract reasoning--the Dominoes Test and the Card Pattern Test--involved stimuli that might be expected to be familiar to all personnel.

1. Memory for Numbers (MEM). A digit-span test measure of immediate memory administered by tape recording. It consists of 21 separate number series, each containing from four to ten digits. A series of numbers is read aloud and after a pause of 8 seconds, subject records numbers remembered and then location. MEM is patterned after a similar test by A. R. Jensen (1964), which has been reported to have substantial validities for school performance and to be largely uncorrelated with IQ. In addition, Jensen reported that digit span tests are relatively culture fair.

Correct recall of digits together with their exact locations were considered in scoring MEM. The following scores were used.

a. Total Correct--The total number of correct recalls.

b. Weighted Total Correct--The sum of the number of correct digits in each span weighted by the total number of digits in the span. Thus, correct answers in a four-digit span would count twice as much as those in a two-digit span.

c. First Error Span--Number of digits in the smallest span with an error.

2. Readministration of the Armed Forces Qualification Test (AFQT). The AFQT contains four 25-item subtests covering (1) vocabulary and verbal reasoning, (2) knowledge of tools and mechanical parts, (3) arithmetic reasoning, and (4) spatial reasoning.

It was suspected that AFQT administration varied from one AFEES to another and that the AFQT scores of some IVs might not reflect their true ability because of test anxiety. Therefore, to provide a basis for comparison, the AFQT was readministered under controlled conditions, providing the following predictors:

a. Vocab./Verbal Subtest Score (total rights).

b. Tool/Mech. Knowledge Subtest Score (total rights).

c. Arithmetic Reasoning Subtest Score (total rights).

d. Spatial Reasoning Subtest Score (total rights).

e. AFQT Total Retest Score (expressed in centiles).

3. Listening Skills I. A tape-recorded test consisting of 38 5-choice, orally-presented problems. The problems involved identifying geometric figures and number series, following directions to find designated locations on a map grid, or visualizing the movement of arrows around a dial. The score was determined by subtracting one-fifth of the "wrongs" from the "rights."

This test was included in the battery because orally administered tests containing simple reasoning items appear to have advantages for low mental level personnel. Since the tests make low demands on academic skill, they appear to generate less test anxiety than paper-and-pencil tests. At the same time, they measure abilities that are important for many relatively unskilled jobs.

4. Listening Skills II. A 37-item, aurally administered, multiple-choice test similar to Listening Skills I except that problems involved direction-following, map-reading, number-series, form-comparison, and dial-reading skills.

5. Dominoes Test (DOM). An 88-item free-answer test involving pictorial problems of dominoes arranged in patterns and series. Subject must note similarities and differences in patterns formed by dominoes. The tightly-timed power test was adapted from the D-48 test developed in Europe during World War II. Gough and Domino (1963) found the D-48 to be highly saturated with "g" and to be entirely nonverbal. They maintained that domino problems draw on a "background of experience that is found in nearly all literate societies" and that the test represents "a promising cross-cultural measure of ability." The score consisted of rights minus wrongs.

6. Card Pattern. A 120-item power test that uses line drawings of playing cards (clubs, diamonds, hearts, and spades) arranged in patterns and series to measure ability to recognize relationships and patterns. The test was an adaptation of the Dalaba Concept Formation Test.[1] Results of research performed in other organizations showed that Dalaba test scores for Blacks had relatively high validity and little of the deficit in validity usually found in the scores of Blacks versus Whites. The score consisted of total rights.

7. Coding. A 120-item speed test in which each item must be coded from a choice of five symbols. This type of test was mentioned by Ghiselli (1966, p. 16) as an effective predictor of on-job performance. The score consisted of total rights.

8. Mechanical Principles. A 60-item test containing pictorial representations of everyday problems that can be solved by applying physical and mechanical principles and relationships. This test was developed because previous research (Ghiselli, 1966, pp. 17-18) had shown that tests of mechnical principles were effective for predicting on-job performance. The score consisted of total rights.

---

[1]Dalaba, E., personal communication, 1968.

Motivation and Perceptual/Motor Skills

1. Hand Skills (HST). The answer sheet for this test comprises a series of numbered boxes. Subjects are tasked to make as many tally marks (viz., 𝍸) as possible in four separately timed sessions. A subject's normal rate of tallying (base rate) is determined from Part 1, which is a 2-minute session described as a practice period to conceal its true purpose. At the beginning of each of the next three parts of the test, a "passing" score is announced. The number of tallies completed beyond the "passing" score is considered a measure of motivation.

The test was developed on the assumption that it measures motivation for carrying out routine or boring tasks--types of jobs frequently handled by IVs. Kipnis (1962) reported that it had significant validities with on-job performance for low-mental ability personnel in Aviation Machinist's Mate, Radioman, and Nuclear Power ratings. The following variables were scored:

a. Part 3 Minus Part 1--Number of tallies on Part 3 minus the number of tallies on Part 1 (base rate).

b. Part 4 Minus Part 1--Total number of tallies on Part 4 minus the total number of tallies in the subject's base rate.

c. Total tallies.

2. Manual Speed (MST). A test resembling the HST in format and presentation. The MST was developed to eliminate the hand scoring which is required for the HST. Like the HST, it has a section for determining base rates, described as a practice period, and timed parts for which passing scores are announced. However, the responses on the MST are not recorded as tallies, but as blackened response circles on an Optical Mark Reader answer sheet. Thus, the MST can be scored electronically.

The following scoring variables were used:

a. Part 3 Minus Part 1--Number of circles blackened in Part 3 minus the number blackened in Part 1 (base rate).

b. Total Score--Total circles blackened in the three parts of the test.

3. Maze Test (MAZ). A group-administered speed test of 60 true/false items in 12 large mazes, each having five pathways leading toward a center goal box. The task is to determine whether or not each pathway is clear to the goal or is blocked.

MAZ is patterned after the Porteus Maze Test, an instrument that has been used in a number of studies involving the illiterate, delinquent, and mentally retarded. The Porteus has been reported favorably as a culture-fair measure of native ability by Porteus (1965) and Jensen (1961). Since the individual scoring required for the Porteus would prevent its being used operationally, the machine-scorable MAZ was developed. The score for MAZ consisted of rights minus wrongs.

4. Word Finding--A 60-item speed test involving matching target words from one column with coded response words in a second column. Matching words are in close proximity in initial items and gradually become further apart as the test proceeds. The subject must find the coded word and copy the symbol following it on an answer sheet. The test is a measure of perceptual speed, an ability which is related to job performance (Curtis, 1971; Cory, 1976; Ghiselli, 1966). The format of the test suggested that it might require greater concentration and less distractibility than the Navy entry subtest used to measure perceptual speed (CLER). The score consisted of total rights.

Vocational Interest

1. Job Check List. A 180-item job preference inventory consisting of descriptions of specific tasks performed by Navy enlisted personnel, such as repairing and adjusting equipment and performing blood tests in a hospital lab. Subjects are asked to indicate whether they like, dislike, or are indifferent to each task. Fifteen scales of occupational interests were derived from a cluster analysis of JCL responses. After a preliminary analysis, scores for the following seven scales were computed and used in the study.

a. Clerical--Keeping records, sorting, operating office machines, typing (21 items).

b. Dangerous Military--Assembling and placing mines, firing torpedos, making parachute jumps, loading ammunition (21 items).

c. Domestic Activities--Cooking, washing, cleaning, sewing (9 items).

d. Communications--Sending and receiving messages by means of radar, radio, sonar, morse code, signal flags (14 items).

e. Verbal Activities--Writing and editing newspaper stories, interviewing, lecturing, stenography (10 items).

f. Graphics--Drafting maps, diagrams, cartoons, illustrations (4 items).

g. Clerical Hand Skills--Typing, handling bank deposits, and withdrawals (3 items).

2. Strong Vocational Interest Blank (SVIB). The SVIB has had a long and distinguished career as an individual guidance tool for vocational counseling. Results with the test over a period of more than 40 years have demonstrated that it has high validity for predicting occupational persistance. Recent studies at this Center (Abrahams & Neumann, 1973; Abrahams, Neumann, & Githens, 1968; Lau & Abrahams, 1971) have found the SVIB to be a useful predictor of retention and performance of enlisted personnel, and have suggested that the SVIB should be evaluated as a predictor of on-job performance. Since the college-level vocabulary of the SVIB presented a potential problem for administering the test to Category IVs, simplified synonyms were prepared for words which were unfamiliar to IVs. SVIB scales that were used as predictors appear below.

a. Business Management (BU)
b Merchandising (MR)
c. Office Practices (OF)
d. Military (MI)
e. Technical Supervisor (TN)
f. Mathematics (MA)
g. Science (SC)
h. Mechanical (MC)
i. Nature (NA)
j. Teaching (TC)
k. Navy Interest Scale (NVR)
l. Achievement (ACH)
m. Liberal-Conservative (L-C)
n. Masculinity-Femininity (M-F)
o. Social Introversion (SI)
p. Occupational Level (OCL)

These variables were scaled as Navy Standard Scores, with means of about 50 and standard deviations of about 10 for large unbiased samples.

3. Performance Index (PI). The PI was originally constructed as a before-after measure of change in maturity of Category IVs occurring during their military service (Edgerton & Sylvester, 1967). However, because of the possibility that maturity might also affect job performance, it was desirable to evaluate the PI as a potential selection and classification test for IVs. For this purpose, a shortened 75-item form of the PI was developed. After preliminary analyses, the following predictors were formed:

a. Personal Maturity--The total number of "true" responses given on a subscale of 25 true/false items on the PI. The areas measured were social perceptions and attitudes, self-confidence, self-actualization, and other aspects of personal maturity.

b. General Maturity--The total number of true responses on the 75 true/false items in the revised PI. Edgerton and Sylvester (1967) consider this type of score a measure of general maturity in young adult males.

Biographical/Attitudinal Data

1. Biographical Information Form (BIF). A questionnaire containing 187 items (with 3- to 5-response alternatives) covering socioeconomic status, family characteristics and living conditions, social experiences, educational background, and preferences and attitudes. From a previous analysis of the BIF for a small subsample of IVs and I-IIIs, Bowser (1974) had developed a scale for predicting reenlistment. Scale scores for the items identified by Bowser were computed by counting the responses to items that distinguished reenlistees from nonreenlistees. In addition, a summary score for socioeconomic status (SES) was computed using weighted responses to seven indices of socio-economic status in the BIF, and a total BIF retention scale was derived. The items from the BIF that were included in the analysis are listed below.

a. Father's level of occupation.

b. Father's level of education.

c. Mother's level of education.

d. Type of neighborhood (rich/poor) in which individual grew up.

e. Did parents live together during most of school years?

f. During the school years, the relative affluence of the family.

g. Number of brothers and sisters who have gone to college.

h. Number of brothers and sisters who dropped out of high school.

i. Educational attainment expected of the individual by his parents or guardians.

j. Number of books read on own time during the last 3 months.

k. Good luck is more important than hard work for success.

l. Every time I try to get ahead, something or somebody stops me.

m. People like me don't have much of a chance to be successful in life.

n. Socioeconomic Status. (Derived scale based on weighted item responses.)

o. Total BIF retention scale. (Derived scale based on 0-1 scoring of responses for 35 items.)

2. Recruit Attitude Questionnaire (RAQ). The RAQ was developed to measure the preferences of enlisted personnel for Navy jobs. A section of the RAQ contains 45 paired-comparison items that provided ipsative measurements of the importance of nine different job characteristics. The following seven scales from the RAQ were used: (a) fringe benefits and retirement, (b) equipment and facilities, (c) patriotism and duty, (d) travel and adventure, (e) pay, (f) prestige, and (g) advancement. The scales are nine-point ipsative scales based on the paired-comparison items in the RAQ. Each total indicates the number of times the category was chosen as more important than its paired category. Thus, the scales constitute rough measures of interest in particular aspects of Navy enlisted service.

3. A Biographical Information Questionnaire (BIQ). The BIQ was designed to assess such factors as cultural background, exposure, and socioeconomic status. A number of scales were derived from the 148 items for use as predictors. Five of them were derived by principle component analysis of responses to the 132 items with scalable responses.

They consisted of the factor scores of the first five principle components identified in the analysis. In addition, scales were derived a priori to contain items that appeared logically to have high interrelationships. All scales from the BIQ are listed below:

a. Factorially Derived:

(1) Activities and Experiences (11% of variance).

(2) Family Possessions (5% of variance).

(3) Electro/Mechanical Interests (4% of variance).

(4) Socioeconomic Status (2% of variance).

(5) Social Leadership (2% of variance).

b. Logically Derived:

(1) Impoverishment of Home Environment--The number of common household items present in the subject's home during his teenage years.

(2) Teenage Activities--Degree of participation in such activities.

(3) Home Activities and Responsibilities--The extent and frequency of home activities and responsibilities.

(4) Mechanical and Technical Experiences--Frequency of performance of the mechanical and technical tasks included in the Activities scale.

(5) Recreation, Sports, and Hobbies--The breadth and intensity of participation in recreational and sports activities.

(6) Literary/Cultural Interests--Degree of participation and interest in literary and cultural activities.

(7) Travel Experiences--Breadth and frequency of travel activities.

(8) Economic and Financial Responsibilities--Breadth and responsibility of experience in economic and financial matters.

(9) Interpersonal Interaction--Level/frequency of such interactions.

(10) Socioeconomic Status--A composite of the weighted elements usually included in SES scales. Major elements were family standard of living, neighborhood, type of home, and the level of education and occupations of parents.

(11) Perceived Social Class--Subject's self-reported social class.

4. Recruit Temperament Survey (RTS). A measure of feelings, emotions, and reactions to experiences. This 105-item (true/false) test was developed to measure neurotic tendencies that would interfere with individual adjustment to living conditions in the Navy, and it was used operationally to screen out probable personality misfits. The questions are phrased so that yes-no answers are required. They are scaled so that the higher an individual's score, the less likely he is to be suitable for Navy service. Studies of the predictive validity of the RTS have found that it correlates significantly with successful adjustment during Boot Camp and throughout the first enlistment, and that it is relatively independent of tests in the enlisted classification battery (Waite & Barnes, undated; Raines, Wittson, & Hunt, 1954).

DISTRIBUTION LIST

Military Assistant for Training and Personnel Technology, Office of the Under Secretary of Defense for Research and Engineering
Director for Acquisition Planning, Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics)
Chief of Naval Operations (OP-115) (2), (OP-11), (OP-987H)
Chief of Naval Material (NMAT 08D2)
Chief of Naval Research (Code 450) (3), (Code 452), (Code 458) (2)
Chief of Information (OI-2252)
Director of Navy Laboratories
Chief of Naval Education and Training (Code-02), (N-2), (N-5)
Chief of Naval Technical Training (Code 017)
Commander, Naval Military Personnel Command (NMPC-013C), (NMPC-001), (NMPC-0481)
Commander, Navy Recruiting Command (Code 20)
Commanding Officer, Office of Naval Research Branch Office, Chicago
Commanding Officer, Recruit Training Command, San Diego
Master Chief Petty Officer of the Force, Naval Education and Training Command (Code 003)
Army Research Institute for the Behavioral and Social Sciences (Reference Service)
Military Enlistment Processing Command (MEPCT-P)
Manpower and Personnel Division, Air Force Human Resources Laboratory, Brooks Air Force Base
Scientific and Technical Information (STINO) Office (AFHRL/TSR), Brooks Air Force Base
Defense Technical Information Center (12)