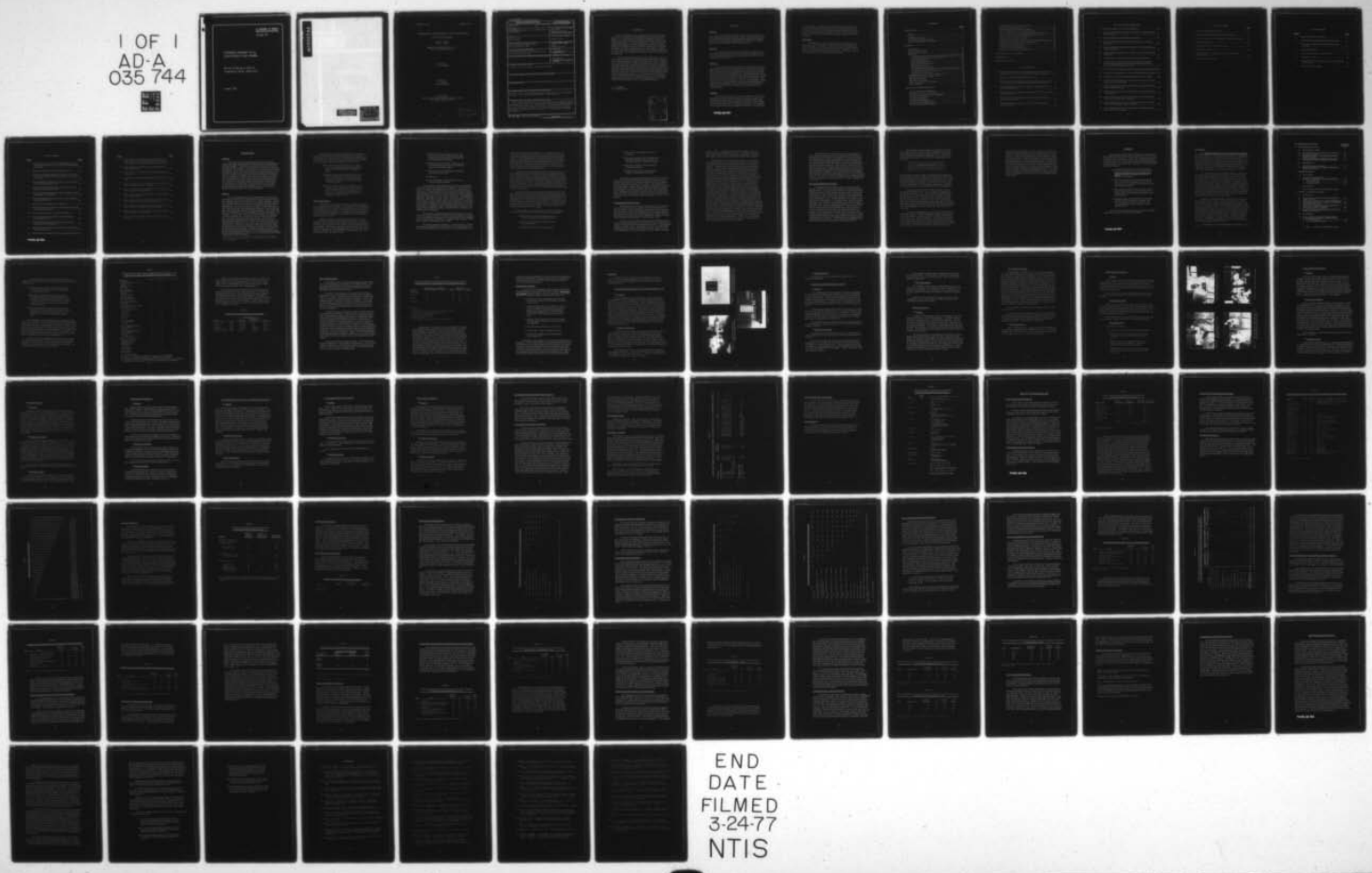
AD-A035 744 APPLIED PSYCHOLOGICAL SERVICES INC WAYNE PA F/G 5/9
EXPERIMENTAL PROCEDURES FOR THE CLASSIFICATION OF NAVAL PERSONN--ETC(U)
JAN 77 A I SIEGEL, J P WIESEN N00123-75-C-1543
UNCLASSIFIED NPRDC-TR-77-3 NL



1 OF 1
AD-A
035 744

END
DATE
FILMED
3-24-77
NTIS

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

AD-A035 744

EXPERIMENTAL PROCEDURES FOR THE CLASSIFICATION OF NAVAL PERSONNEL

APPLIED PSYCHOLOGICAL SERVICES,
INCORPORATED, WAYNE, PENNSYLVANIA

JANUARY 1977

AD A035744

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

DDC
RECEIVED
FEB 18 1977

NPRDC TR 77-3 January 1977

EXPERIMENTAL PROCEDURES FOR THE CLASSIFICATION OF NAVAL PERSONNEL

Arthur I. Siegel
Joel P. Wiesen

Applied Psychological Services, Inc.
Wayne, Pennsylvania 19087

Reviewed by
M. F. Wiskoff

Approved by
J. J. Regan
Technical Director

Prepared for
Navy Personnel Research and Development Center
San Diego, California 92152

DDC
RECEIVED
FEB 18 1977
A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR 77-3	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* EXPERIMENTAL PROCEDURES FOR THE CLASSIFICATION OF NAVAL PERSONNEL		5. TYPE OF REPORT & PERIOD COVERED Final Report May 1975 - May 1976
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Arthur I. Siegel Joel P. Wiesen		8. CONTRACT OR GRANT NUMBER(s) N00123-75-C-1543
9. PERFORMING ORGANIZATION NAME AND ADDRESS Applied Psychological Services, Inc. Wayne, Pennsylvania 19087		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62763N ZF55-521-031-03-01
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		12. REPORT DATE January 1977
		13. NUMBER OF PAGES 79
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		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)*

Assessment Center, Classification, Evaluation, Performance Tests, Work Sample, Simulation, Test Theory, Job Classification, Policy Capturing

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

Two concepts--miniature job learning and evaluation and assessment center methodology--were woven into a technique for evaluating and classifying personnel for technically oriented jobs. The concepts are presented and the resultant evaluative methodology described. Trial work indicated acceptable internal psychometric characteristics and considerable acceptability for the methods and approach.

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

FOREWORD

The overall goal of this research effort was to develop and demonstrate a method for reclassifying General Detail Navy Personnel, i.e., those who did not qualify for assignment to an "A" school on the basis of aptitude scores derived from a conventional classification battery. Recommendations for reclassification were based upon an assessment center approach in which job learning tasks and job sample tests were utilized in determining an individual's ability to perform successfully in six Navy ratings. Results from this work will be employed in future research and development to design, test, and evaluate a classification system for General Detail personnel.

The monitor for this contract was Mr. Paul Foley. Mr. Ted Yellen also contributed by providing practical assistance and advice in the establishment and conduct of the assessment center. Appreciation is also extended to LCDR Gary Johnson and to CPO John Gavel of the Apprenticeship Training Department of the Recruit Training Command, Naval Training Center, San Diego, who performed the necessary arrangements relative to facilities, scheduling, and the like. The major contributions during the test administrative phases of CPO Robert G. Semperger and CPO William Parks of the Apprenticeship Training Department, and Mr. W. Rick Leahy and Dr. Robert M. Voytas, both of Applied Psychological Services, are acknowledged. These persons also made a number of suggestions for increasing the effectiveness and efficiency of the methods employed.

J. J. CLARKIN
Commanding Officer

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. Avail. and/or SPECIAL
A

SUMMARY

Problem

Current Navy Recruit classification methods are based on paper-and-pencil testing techniques. There is no available method for reclassifying persons who are originally assigned to "general service" but who warrant reclassification. The result is often improper and noncost/effective personnel utilization.

Objective

The purpose of this effort was to develop a method for reclassifying Navy personnel who did not qualify for assignment to an "A" school on the basis of aptitude scores derived from a conventional classification battery.

Approach

Two recent evaluative developments--the assessment center approach and an exercise development model which is based on the logic that a person who demonstrates the ability to learn a sample of a job will be able (given appropriate training) to learn the total job--were woven into a classification/reclassification scheme. The scheme is based on "hands on" training procedures and work sample performance evaluative methods. Navy jobs were analyzed, and a set of exercises, based on the two concepts, was developed. These exercises were then incorporated into a traditional management assessment center paradigm and the methods tried on 140 enlisted persons who were previously judged to be not qualified for "A" school training.

Findings

Of the 140 men involved, 76 were considered, as the result of application of the methods employed, to possess capability for adequate performance in one of the following specialties: Postal Clerk, Storekeeper, Yeoman, Signalman, Machinist's Mate, Hospital Corpsman, Engineman, and Electrician's Mate. The reliability of the exercises and of the subsequent classification decisions was

Preceding page blank

very acceptable. Consistent classification policies were followed by the various staff members from whom the classification recommendations evolved. The individuals assessed considered the multiple assessment methods employed to be fairer than and preferable to the usual methods which are hinged to paper-and-pencil evaluative techniques.

Conclusions

The approach to classification/reclassification here developed seems entirely workable, useful, and within acceptable standards of fairness. At least from the points of view investigated, the approach also possesses adequate psychometric properties. The approach can identify persons who seem to possess capability which is not identified by the usual classification methods.

CONTENTS

	Page
INTRODUCTION	1
Problem	1
Purpose	1
Assessment Centers	2
Job Sample Performance Tests	5
Job Learning Approach to Testing	7
METHODS	11
Job Analysis	12
Exercise Development	17
Administrative Methods	19
Exercises	20
Conceptual Integration/Application (Troubleshooting) Exercise	20
Coordinative Speed and Accuracy Exercise	22
Dual Task Exercise	23
Inspection/Sort Exercise	25
Level of Aspiration Exercise	27
Reliability Exercise	28
Social Interactive Exercise	29
Tool and Object Nomenclature, Use, and Recognition Exercise	30
Writing Sample Exercise (Optional)	31
Semistructured Interview	32
Briefing/Debriefing and TCAC Evaluation Form	33
Administrative Procedures for TCAC	33
Individuals Assessed	34
Description of Sample	34
TCAC Staff and Time of Operation	36
Data Considered	36
RESULTS AND INTERPRETATION	39
Career Field Recommendations	39
Predictions of On-the-Job Success	39
Exercise Scores and Dimensional Ratings	41
Dimensional Judgments	41
Interrater Reliability	44
Dimensional Evaluations	46
Interdimensional Relationships	46
Intradimensional Relationships	47

Psychophysical and Motor Dimension 49
Social/Motivational Dimension 49
Policy Capturing of Dimensional Ratings 52
Policy in Rating Men on Learning Dimension 53
Policy in Rating Men on Psychophysical-Motor Dimension 56
Policy in Rating Men on Social/Motivational Dimension 57
TCAC Career Field Recommendations 58
Prediction of On-the-Job Success 60
Group Policy For Machinist's Mate and General Rates Subsamples .. 61
Group Policy With Respect to the Total Sample 63
Comparison of Navy and TCAC Scores 65
Views of Individuals Assessed 67
Reasons for Fairness Evaluation 68
Opinions Concerning Specific Exercises 69

DISCUSSION AND CONCLUSIONS 71

REFERENCES 75

DISTRIBUTION LIST 81

LIST OF TABLES

1. Requirements of Various Navy Ratings Along the Learning, Psychophysical and Motor, and Social/Motivational Dimensions 15

2. Ratings to be Recommended and Requirements 16

3. Structure-of-Intellect Analysis of the Exercises Used to Assess the Learning and the Psychophysical and Motor Dimensions 18

4. Mean and Standard Deviation of Sample on Navy Related Scores and Biographical Items 35

5. Number and Source of Data Scores Recorded for Each Individual Assessed/Classified 37

6. Career Field Recommendations and Corresponding Mean Predictions of Job Success 40

7. Mean and Standard Deviation (S. D.) of TCAC Judgment and Exercise Scores 42

LIST OF TABLES (CONTINUED)

8. Intercorrelations Among Various Measures 43

9. Interrater Reliability (Intraclass Correlation) of Selected Center Exercise Scores..... 45

10. Intercorrelations Among Dimensional Ratings 46

11. Intercorrelations Among TCAC Scores Relevant to Learning Dimension..... 48

12. Intercorrelations Among Scores Relevant to Psychophysical and Motor Dimension..... 50

13. Intercorrelations Among Scores Relevant to Social/Motivational Dimension..... 51

14. Group Policy for Rating Men on the Learning Dimension..... 54

15. Relative (Order of Regression Entry) Importance of Center Scores Most Weighted in the Policies of Each Administrator and of the Administrators as a Group..... 55

16. Group Policy for Rating Men on the Psychophysical-Motor Dimension 57

17. Group Policy for Rating Men on the Social/Motivational Dimension.. 58

18. Comparison of Actual and Statistical Classifications of Men..... 60

19. Policy Capturing of Level of Job Success in the Fleet by Score--Machinist's Mate Subsample..... 61

20. Policy Capturing of Level of Job Success in the Fleet by Score--General Rates Subsample..... 62

21. Policy Capturing of Level of Job Success in the Fleet by Score--Total Sample 64

22. Policy Capturing of Level of Job Success in the Fleet by Navy Test Score--General Rates Subsample..... 66

23. Policy Capturing of Level of Job Success in the Fleet by Navy Test Score--Machinist's Mate Subsample 66

24. Policy Capturing of Level of Job Success in the Fleet by Navy Test Scores--Total Sample 67

LIST OF FIGURES

		Page
1.	Sample NAVPERS 18068C page	13
2.	Conceptual Integration/Application administration	21
3.	Tools and equipment for Coordinative Speed and Accuracy	21
4.	Dual Task equipment	21
5.	Inspection/Sort exercise administration	26
6.	Reliability exercise	26
7.	Tool and Object Nomenclature, Use, and Recognition administration	26
8.	Post exercise debriefing	26

LIST OF FIGURES

Figure		Page
1	Sample NAVPERS 18068C page	13
2	Conceptual Integration/Applicator administrator	21
3	Tools and equipment for Coordinative Speed and Accuracy	21
4	Dual Task equipment	21
5	Inspection/Sort exercise administration	26
6	Reliability exercise	26
7	Tool and Object Nomenclature, Use, and Recognition administration	26
8	Post exercise debriefing	26

LIST OF TABLES

Table		Page
1	Requirements of Various Navy Ratings Along the Learning Psychophysical and Motor, and Social/Motivational Dimensions	15
2	Ratings to be Recommended and Requirements	16
3	Structure-of-Intellect Analysis of the Exercises Used to Assess the Learning and the Psychophysical and Motor Dimensions	18
4	Mean and Standard Deviation of Sample on Navy Related Scores and Biographical Items	35
5	Number and Source of Data Scores Recorded for Each Individual Assessed/Classified	37
6	Career Field Recommendations and Corresponding Mean Predictions of Job Success	40
7	Mean and Standard Deviations (S. D.) of TCAC Judgment and Exercise Scores	42
8	Intercorrelations Among Various Measures	43
9	Interrater Reliability (Intraclass Correlation) of Selected Center Exercise Scores	45
10	Intercorrelations Among Dimensional Ratings	46
11	Intercorrelations Among TCAC Scores Relevant to Learning Dimension	48
12	Intercorrelations Among Scores Relevant to Psychophysical and Motor Dimension	50
13	Intercorrelations Among Scores Relevant to Social/Motivational Dimension	51
14	Group Policy for Rating Men on the Learning Dimension	54

Preceding page blank

Table		Page
15	Relative (Order of Regression Entry) Importance of Center Scores Most Weighted in the Policies of Each Administrator and of the Administrators as a Group	55
16	Group Policy for Rating Men on the Psychophysical-Motor Dimension	57
17	Group Policy for Rating Men on the Social/Motivational Dimension	58
18	Comparison of Actual and Statistical Classifications of Men	60
19	Policy Capturing of Level of Job Success in the Fleet by Score--Machinist's Mate Subsample	61
20	Policy Capturing of Level of Job Success in the Fleet by Score--General Rates Subsample	62
21	Policy Capturing of Level of Job Success in the Fleet by Score--Total Sample	64
22	Policy Capturing of Level of Job Success in the Fleet by Navy Test Score--General Rates Subsample	66
23	Policy Capturing of Level of Job Success in the Fleet by Navy Test Score--Machinist's Mate Subsample	66
24	Policy Capturing of Level of Job Success in the Fleet by Navy Test Scores--Total Sample	67
25	Example Reasoning for Perceiving the TCAC Exercises as Fair	68

INTRODUCTION

Problem

The Navy has traditionally based classification decisions largely upon paper and pencil tests which have been validated against "A" school grades. The problem is that direct measures of an individual's ability to perform the requirements of a job are not directly assessed. Accordingly, a large number of recruits who are unable to qualify for formal "A" school training are assigned directly to the Fleet where they must rely upon on-the-job training and personal initiative to compete successfully for promotion with their "A" school trained counterparts. Little is known about the potential of general service personnel to perform in a formal school setting because of their inability to satisfy the entrance requirements as determined by test results. In the present effort representative performance job samples were developed to investigate their utility as supplemental measures in assessing "general service" personnel for reclassification and possible "A" school assignment.

Purpose

The overall goal of the present program was to develop and demonstrate a method for reclassification of Navy personnel who: (a) are originally considered unsuitable for "A" school training, or (b) desire reclassification because they are not satisfied with their original classification. The need for improved methods for personnel reassessment stems from the fact that, at present, the Navy possesses no fair, objective, valid, and standardized technique for reassessing the enlisted personnel after they are initially classified. At the time of the study, the classification method depended on the use of the Navy's "basic battery." This battery consists of a general aptitude test (GCT), a mechanical aptitude test, an arithmetic test, an electronic aptitude test, a shop practices test, and a clerical test.[1] On the basis of scores on these instruments plus interest information, a recruit is classified and assigned to an "A" school or directly to the Fleet. Misclassifications, expecially in the case of those recruits sent directly to the Fleet, are costly to the Navy and to the general society. Such misclassified persons are apt to advance only slowly in the Fleet and are more apt, as a consequence, to become disciplinary problems. Moreover, and of greater importance, due to misclassification, a sailor may be classified/assigned so that he contributes less than his full potential.

[1] As of January 1976 the Armed Services Vocational Aptitude Battery (ASVAB) has replaced the Basic Test Battery (BTB) as the classification instrument.

The method developed and evaluated here is called the Technical Classification Assessment Center (TCAC) approach. It emphasized exercises which allow the individual to learn and perform tasks that simulate likely on-the-job learning and performance. The basic concepts underlying this method were:

- The individual deals with any task by applying in an integrated fashion his total complement of aptitudes and abilities (Rundquist, 1969).

- Results from exercises which allow the individual to demonstrate how well he can learn aspects of a job can provide a basis for predicting performance on the whole job (Siegel & Bergman, 1975).

- The assessment center performance test approach, in which several test administrators representing possible diverse points of view observe an individual as he performs job related tasks, has promise for technical level jobs (Bray & Moses, 1972).

Assessment Centers

Current use of assessment centers is limited almost entirely to managerial level personnel. The extension of the assessment center approach to technical level jobs represents an elaboration of the assessment center concept, as originally developed. However, the extension of the assessment center approach to nonmanagerial jobs was previously suggested as a possible area of investigation by Bray and Moses (1972).

Several reviews of current assessment center practices are available (MacKinnon, 1975; Howard, 1974; Huck, 1973; Bray & Grant, 1966). The assessment center approach treats each individual as a whole person rather than as a sum of specific abilities and aptitudes. The typical assessment center runs for one to three days and involves two to five administrators (assessors) and six assessees. The goal is to predict future managerial success. Typical measurements and tasks include:

- Inbasket tests--memoranda, letters, and notes are presented, and the person being assessed indicates (usually in writing) how he would handle each of these

- Group discussion--usually a leaderless discussion possessing the goal of formulating a solution to a specific problem

- Objective tests--including paper and pencil aptitude and ability tests

- Interview

- Other techniques--such as projective tests and role playing situations

The assessment center approach represents a consolidation of clinical, gestalt, and behavioristic evaluation. The clinical approach is evident in the judgments involved in formulating assessment center recommendations based on data from and observation of performance of the above and similar tasks on variables such as: aggressiveness, energy level, interpersonal contact, self confidence, sensitivity, dependence on others, work motivation, and adaptability (Hinrichs, 1969; Bray & Grant, 1966). The gestalt aspect is seen in the assessment center's final rating of each individual as a whole person (in terms of whether or not or how well the person will succeed) rather than a rating based on a statistically weighted sum of various scores (Bray & Campbell, 1968). The behavioristic influence is reflected in the emphasis on rating of actual behaviors in (largely simulated) real life situations.

There is also an apparent psychometric influence. The psychometric influence is seen in the fairly numerous attempts to measure the reliability and the validity of the assessment center judgments (e.g., Bray & Grant, 1966; Bray & Campbell, 1968; Hinrichs, 1969; and Greenwood & McNamara, 1967, 1969).

The present program represents, to our knowledge, the first extension of the assessment center approach to technical jobs (such as the Navy machinist's mate, postal clerk, and storekeeper). All

of the major aspects of the assessment center approach were preserved in the current TCAC. However, the evaluative instruments often used in assessment centers are designed for managerial positions. Tests of this nature (e.g., inbasket tests) are not appropriate for technical level jobs. In the present TCAC, such assessment tools were replaced by exercises related to technical jobs. Following usual assessment center philosophy, the assessee is required, in dealing with these exercises, to draw on all of his abilities and aptitudes in an integrated fashion.

The validity of the assessment center approach has been usually evaluated in terms of prediction of future managerial success. In this essential respect, the assessment center approach seems to work. Predictive validity correlation coefficients of .40 and higher are not uncommon in the literature (MacKinnon, 1975; Ash & Kroeker, 1975; Mitchel, 1975; Howard, 1974; Bray & Moses, 1972; and Wollowick & McNamara, 1969). Prior to the availability of validity data, an assessment center may be evaluated on the basis of descriptive statistics relative to the assessment situations employed and the judgments of the assessors.

Evaluation of an assessment center should also include consideration of several aspects of the practical operation of the center. The assessment center should be easily administered and should not unduly disrupt the ongoing operation of the parent organization. The assessment experience should be a profitable one for the assessees, and the experience should possess face validity. The public relations function that face validity serves (APA, 1974) is becoming more important as the general public becomes more sophisticated and more wary concerning psychological testing.

As the result of a recent Nebraska Court Decision (Anonymous, 1976), several other important criteria have emerged:

- Administrators must be adequately trained in how to conduct the assessment center.
- The assessment materials should be based on a job analysis.
- The exercises used must be appropriate.

- There must be an adequate number of exercises.
- Assessors should not know candidates before the assessment and assessors and candidates should be randomly assigned.
- The center should be administered in a professional manner.
- Assessor effectiveness and reliability should be continuously monitored.

In planning and implementing any testing program, consideration of certain ethical questions is required. Basic ethical guidelines for assessment center operations have been developed and are summarized by Moses et al. (1975). They discuss five areas that require consideration: organizational support for assessment operations, assessor training, informed consent on the part of participants, use of assessment center data, and validation issues. Although this statement on ethical principles was not available at the onset of this project, these areas were considered and in all cases the recommended minimum standards were met or surpassed.

Job Sample Performance Tests

Performance tests in general and particularly the job sample variety have become the object of increased recent interest for two reasons. First, performance tests may yield an improvement over the modest levels of prediction of job success which result from the use of the paper-and-pencil tests. Second, job sample tests seem to conform to the job relatedness criterion of recent court decisions (Ash & Kroeker, 1975; O'Leary, 1973).

The idea of using a sample of actual work behavior as a predictor of future job success is a relatively old one. The use of job sample tests in clerical settings dates back at least to 1933 (Albright, Glennon, & Smith, 1963). As long ago as 1947, a text describing job sample performance tests described in detail most, if not all, of the techniques in use today (Adkins, Primoff, McAdoo, Bridges,

& Forer, 1947). Through the years these techniques have been applied often in both military (e.g., Greer, Pearson, & Havron, 1957; Jensen, Hill, Siegel, & Courtney, 1954; Siegel & Courtney, 1953) and civilian (e.g., Campion, 1972; Tiffin, 1952) settings.

Performance tests are usually considered in contrast to paper and pencil tests. One class of performance tests attempts to measure mechanical or psychomotor abilities or aptitudes in isolation. Tests of this type are of little interest in the present context. A second class of performance test, the job (or work) sample performance test, requires the individual to perform a task which is similar or identical to one performed on the job. Two different philosophies underlie the use of these two types of performance tests. The tests of specific psychomotor abilities are based on a specific abilities philosophy. This philosophy holds that final job performance can be predicted based on measures of the independent unitary abilities required by the job. Such specific abilities may include: choice reaction time, wrist-finger speed, manual dexterity, and speed of limb movement (McCormick & Tiffin, 1975). The tests that consist of job samples are based on an approach which usually emphasizes the overall capabilities of a person. Each person is assumed to bring to a task a different complement of abilities. These are applied to the performance of the task in unique ways. Two people may perform well based on two different complements of abilities. Rundquist (1969) made this point. Referring to aptitude scores or actual job performance, he wrote, "...phenotypic score similarities are no guarantee of similarities in the basic processes involved." Rundquist also suggested that the failure to consider the dynamic interrelationships of abilities in each person may have limited current methods for developing predictors to their current low levels of validity. It is likely, both from the perception of individuals and in reality, that individuals deal with real life situations by applying their resources as adaptively as possible. This may differ from person to person; actual on-the-job performance may not be based on the same ability or combination of abilities for every person (Rundquist, 1969). Thus, job performance may not be well predicted by one aptitude or ability score or even by a linear combination of such scores.

The performance oriented tests of the present program which simulate real on-the-job activities allow the individual to demonstrate how effectively he can apply his complement of aptitudes and skills to complex situations akin to those encountered in real life by Navy enlisted personnel. In using this approach, one tests the ability of the individual to perform an actual or simulated sample of the actual job. There is some evidence that this approach is valid for technical level jobs (Campion, 1972; McCormick & Tiffin, 1975). Campion (1972) compared job sample tests with several paper and pencil tests (including tests of mechanical comprehension, intelligence, and numerical and clerical ability). Correlations between job sample predictors and criterion variables were .66, .42, and .46, while correlations between paper-and-pencil test scores and the criterion variables were low, ranging from -.23 to .08.

Job Learning Approach to Testing

An approach to performance tests, recently developed through a systematic set of studies performed at Applied Psychological Services, is based on the idea that if a man demonstrates the ability to learn a representative sample of a job, then he can be expected to be able to learn the whole job (Siegel & Bergman, 1975; Siegel & Leahy, 1974; Siegel, Bergman, & Lambert, 1973; Siegel & Bergman, 1972). In this approach, each evaluative situation involves two phases: (1) a training phase, and (2) an evaluation phase. An example of this approach, described by Siegel & Bergman (1975), includes a detailed description of how to start up and shut down a motor and pump apparatus. This procedure involves 33 steps including several safety precautions. Each examinee is then given an opportunity for supervised "hands on" practice of this procedure. After this practice session, a performance test is administered. Other tests of this type described by Siegel and Bergman include: tool identification and use, gasket cutting and meter reading, troubleshooting on a pumping system, and assembly of a gate valve. These tests were developed to assess promise in the Navy machinist's mate career field.

This approach to performance testing has been termed the miniature job training and evaluation approach. An earlier related approach was used by Lawshe and Tiffin in the development of the Purdue Mechanical Adaptability Test. This test, described by McCormick & Tiffin (1975), was developed on the basis that:

> ...there was reason to believe from a previous study that, other things being equal (emphasis added), those persons who have most profited in knowledge from previous mechanical experiences may do better on mechanical jobs than those persons who have not so profited (pp. 143-144).

In concordance with this, Siegel and Bergman (1975) suggested that the miniature task training part of their approach gives all the individuals tested a fair chance to do well on the test in a way that a performance test without a learning phase does not. The effects of differences in exposure to mechanical devices in general and specifically to the job sample considered are thought to be controlled due to the equalizing effect of the training phase.

Siegel and Bergman (1975) constructed their miniature job training and aptitude tests so as to minimize emphasis on the ability to read and write. Ash and Kroeker (1975) cautioned against the use of tests which have a higher reading level than that required by the job. Such tests may be inherently biased against certain classes of people who, for reasons associated with their culture or socioeconomic class, have had less formal education, or less successful formal education, than other classes. For a motivated person, formal education, which yields a facility with written English, may have little to do with job success in many jobs.

The job learning approach as used with low aptitude Navy enlisted personnel by Siegel and Bergman (1972) was successful in several respects. Predictive validity multiple correlation coefficients of .15 to .46 were obtained. This compares favorably with the predictive validity of the Navy test scores in this situation. Siegel and Leahy (1974) extended these findings with reports of subjective evaluations of the sailors selected with the new miniature training and evaluation method. Supervisors of the men

recommended indicated that, if given a choice, in 79 percent of the cases they would choose the recommended man as a subordinate. Siegel and Leahy also reported that despite careful analysis, no race related differences in performance on these tests were detected. Siegel and Bergman (1975) concluded that this type of test has merit and promise in three areas: (a) predicting performance of "low aptitude" applicants, (b) extension to other less mechanical jobs, and (c) valid testing which is also fair to minority groups. The present study adapted several of the miniature training and evaluation tests developed by Siegel and his associates at Applied Psychological Services. The approach is extended in the present program to other technical level jobs, including some jobs that are low on mechanical ability requirements.

METHODS

The overall purpose of the present program was to establish a classification/reclassification technique for non "A" school listed personnel in the Navy. To this end, the assessment center concept was melded with the miniature job learning and aptitude test approach to yield an articulated scheme possessing the merits of both concepts. To achieve this combination, several sequential steps were required:

a. Job analysis--performance of a three dimensional conceptual analysis of NAVPERS 18068C (Manual of Qualifications for Advancement in Rating) and a choice of target Navy career fields for consideration.

b. Exercise development--development of the specific exercises to be included in the TCAC on the basis of the job analysis and Guilford's Structure-of-Intellect model.

c. Administrative methods--development of a complete administrator's manual, training the administrators, and development of decision making procedures.

d. Evaluation plans--development of a systematic procedure for evaluating the adequacy of the TCAC including the exercises, per se, the recommendations, and the reliability of the administrators.

Each of these developments, along with their rationale, is described in subsequent sections of this chapter.

Preceding page blank

Job Analysis

The Manual of Qualifications for Advancement in Ratings (NAVPERS 18068C) was taken as the basic source of information for the job analysis. This Manual contains detailed, updated lists of the tasks performed by personnel in all Navy specialties. The tasks required for a promotion to the E4 level were chosen for analysis as this level is usually reached within the original enlistment period (four years) of typical enlisted Navy personnel. The importance of a close correspondence between the level of the individuals tested and the level of jobs considered in validation studies was emphasized by the Supreme Court (Lifter, 1976).

The task statements in NAVPERS 18068C are different in one respect from many other job analyses. The requirements given in NAVPERS 18068C apply to all enlisted men in the specified rating. The actual jobs men in these ratings perform may vary from ship to ship and from assignment to assignment. However, all men in a given rating are expected to be able to perform the tasks listed for that rating. The use of NAVPERS 18068C was dictated by the practical impossibility of developing a workable system with considerably finer job discriminations than those already in use by the Navy (with nine levels of each of over 80 jobs). The use of NAVPERS 18068C had the added virtue of keeping the present study in step with the most current Navy statement of requirements of the various rates and ratings.

Analysis of NAVPERS 18068C identified three dimensions which appeared to be relevant to all Navy ratings: (a) learning, (b) psychophysical and motor, and (c) social-motivational. This structure, in a sense, parallels the classification of aptitude tests developed by Ghiselli (1966, 1973). Ghiselli classified the various types of aptitude tests into five categories: (a) intellectual abilities, (b) spatial and mechanical abilities, (c) perceptual accuracy, (d) motor abilities, and (e) personality traits. The dimensionalization used in the present study effectively consolidates Ghiselli's three middle categories into one category. These three dimensions were, accordingly, considered to cover comprehensively most skills, abilities, and aptitudes requisite for on-the-job success for enlisted personnel in the Navy.

A sample page from NAVPERS 18068C is presented in Figure 1.

QUALIFICATIONS FOR ADVANCEMENT		Required for Advancement to YN
B.	OFFICE PROCEDURES - Continued	
1.00	Practical Factors - Continued	
.04	Meet enlisted, officer, and civilian personnel as a receptionist	E-4
.05	Speak clearly and distinctly	E-4
.08	Demonstrate courtesy, tact, and good judgment when handling telephone communications	E-4
.40	Organize and maintain files, including files of directives and correspondence, and cross reference sheets and tickler files	E-5
2.00	Knowledge Factors	
.01	Proper English grammar and punctuation	E-4
.40	Regulations concerning disposal, stowage, and transmission of obsolete files and records	E-5
C.	EQUIPMENT OPERATION	
1.00	Practical Factors	
.03	Operate office duplicating equipment	E-4
.39	Typewrite for 5 minutes: (See Performance Test Instructions)	
	a. At 30 words per minute	E-4
	b. At 40 words per minute	E-5
	c. At 50 words per minute	E-6
2.00	Knowledge Factors	
	None in addition to those implied in the above Practical Factors.	
D.	REPORTS, PUBLICATIONS, AND RECORDS	
1.00	Practical Factors	
.01	Transcribe officer message orders, using current instructions	E-4
.02	Maintain publications	E-4
.40	Maintain officer personnel diaries in accordance with Manpower and Personnel Management Information System Manual (MAPMISMAN)	E-5
.41	Maintain officer service records	E-5
.42	Verify the Officer Distribution Control Report (ODCR) and other related reports as required by MAPMISMAN	E-5
.43	Prepare Manpower Authorization Requests	E-5
.60	Supervise the procurement, maintenance, stowage, issuance, and custody of official publications	E-6
2.00	Knowledge Factors	
.01	General content and use of standard publications pertaining to personnel and general administration, including the following:	
	a. U.S. Navy Regulations	E-4
	b. Bureau of Naval Personnel Manual	E-4

Figure 1. Sample NAVPERS 18068C page.

A working definition of each of these dimensions was developed in order to evaluate the various Navy career fields. These definitions are:

- Learning--requirements for acquiring complex knowledges and skills and for integrating these into purposeful activity.

- Psychophysical and motor--requirements for fine discriminations, complex psychomotor coordination, attentiveness, and precision, and for integrating these into a rapid response possibility under stressful and/or time limited and time critical conditions.

- Social/motivational--requirements for demonstrating such characteristics as initiative, leadership behavior, social interactive skill, and personal characteristics consistent with mission goals.

Two independent raters evaluated the knowledge and skill requirements of the general ratings described in NAVPERS 18068C (with some exceptions, such as the aviation group). Based on consideration of the skills and knowledges required for promotion to the E4 level, each judge rated each Navy rating relative to each of the three conceptual categories. The judgments for each category were made on a four level scale rating from "none of this factor is required for adequate job performance" through "a great deal of this factor is required for adequate job performance."

After the two raters independently completed their ratings, any differences between the two were discussed. Based on consideration of the details for each rating, as given in NAVPERS 18068C, any disagreements in the evaluations were resolved. The resulting final evaluations are given in Table 1.

Table 1

Requirements of Various Navy Ratings Along the Learning, Psychophysical and Motor, and Social/Motivational Dimensions

	Learning	*Psychophysical/Motor*	*Social/Motivational*
GROUP I DECK			
Boatswain's Mate (BM)	2	2	3
Electronics Warfare Technician (EW)	3	2	2
Quartermaster (QM)	3	2	2
Operations Specialist (OS)	3	2	2
Signalman (SM)	2	2	1
Sonar Technician (ST)	2	3	2
Master at Arms	2	1	3
GROUP II ORDINANCE			
Fire Control Technician (FT)	3	2	2
Gunner's Mate (GM)	2	2	2
Mineman (MN)	2	2	2
Missile Technician (MT)	2	2	2
Torpedoman's Mate (TM)	3	2	2
GROUP III ELECTRONICS			
Data Systems Technician (DS)	3	2	2
Electronics Technician (ET)	3	2	1
GROUP IV PRECISION EQUIPMENT			
Instrumentman (IM)	2	3	1
Opticalman (OM)	2	3	1
GROUP V ADMINISTRATIVE AND CLERICAL			
Commisaryman (CS)	2	1	2
Communications Technician (CT)	4	1	1
Data Processing Technician (DP)	2	1	1
Disbursing Clerk (DK)	3	1	1
Journalist (JO)	3	1	3
Personnelman (PN)	2	1	2
Postal Clerk (PC)	2	1	1
Radioman (RM)	3	2	1
Ship's Serviceman (SH)	3	1	2
Storekeeper (SK)	2	1	2
Yeoman (YN)	2	1	3
Steward (SD)	2	1	1
GROUP VI MISCELLANEOUS			
Illustrator Draftsman (DM)	2	1	1
Lithographer (LI)	3	1	1
Musician (MU)	2	2	1
GROUP VII ENGINEERING AND HULL			
Boiler Technician (BT)	2	1	1
Electrician's Mate (EM)	3	2	2
Engineman (EN)	3	3	2
Hull Maintenance Technician (HT)	2	2	1
Interior Communications Electrician (IC)	3	2	1
Machinery Repairman (MR)	3	3	1
Machinist's Mate (MM)	2	2	2
Molder (ML)	2	2	1
Patternmaker (PM)	3	1	1
Gas Turbine System Technician (GS)	3	2	2
GROUP VIII CONSTRUCTION			
Builder (BU)	2	2	1
Construction Electrician (CE)	3	2	1
Construction Mechanic (CM)	1	1	1
Engineering Aid (EA)	3	1	1
Equipment Operator (EO)	2	2	1
Steel Worder (SW)	2	3	1
Utilitiesman (UT)	2	2	1
GROUP X MEDICAL			
Hospital Corpsman (HM)	3	2	3
GROUP XI DENTAL			
Dental Technician (DT)	3	2(3?)	3

Key to scale values:

1= little of this dimension required for adequate job performance

2= some of this dimension required for adequate job performance

3= considerable amount of this dimension required for adequate job performance

4= a great deal of this dimension required for adequate job performance

Based on the evaluations presented in Table 1, target ratings (ratings to be considered by the TCAC) were chosen that: (1) reflect a wide range of requisite abilities, and (2) involve a wide range of duties. These goals were established in order to allow assessment of the generality of the TCAC approach.

Two further goals were established in selecting target ratings: (1) the number of ratings chosen was kept to a minimum so that the number of cases in each rating would be large enough for meaningful data analyses, and (2) the prestige of the ratings, in terms of Navy life and the value of the skills of the ratings in terms of post Navy career plans, were kept as high as possible. The ratings selected to be used as target ratings (or career fields) by the TCAC, along with the requisite levels on each of the three dimensions, are given in Table 2.

Table 2

Ratings to be Recommended and Requirements

		Requirements		
Rating		*Learning*	*Psychophysical and Motor*	*Social*
Postal Clerk	(PC)	moderate	low	low
Storekeeper	(SK)	moderate	low	moderate
Yeoman	(YN)	moderate	low	high
Signalman	(SM)	moderate	moderate	low
Machinist's Mate	(MM)	moderate	moderate	moderate
Hospital Corpsman	(HM)	high	moderate	high

Exercise Development

The design, form, and content of the exercises included in the TCAC was guided by the analysis of NAVSHIPS 18068C, which resulted in the identification of the three dimensions: learning psychophysical and motor, and social/motivational, and by the subsequent selection of target ratings.

However, it seemed clear, on the basis of the analytic method used and on the literature, that each of the three dimensions is not, in fact, unidimensional. Accordingly, an effort was made to assure that the exercise battery sampled as many of the individual aspects of these dimensions as possible. To this end, an empirically derived theory concerning the basic factors of intellective function (Guilford & Hoepfner, 1971) was used to help guide the design of the individual exercises and to help assure a balanced test battery.

Guilford's Structure-of-Intellect model of intellective functioning (Guilford & Hoepfner, 1971) was used as a guide in the design and development of the exercises used to evaluate abilities in the learning and the psychophysical and motor dimensions. Exercises were selected and developed so that each combination of Guilford's intellective operations (cognition, memory, and evaluation) and intellective contents (figural, symbolic and semantic) would be required by at least one test. While the other parameters of Guilford's theory seem to be sampled by the exercises developed, emphasis in this developmental stage was placed on the parameters mentioned. The tests developed to assess the learning and the psychophysical and motor dimensions, together with the combination of the Guilford parameters sampled by these tests, are given in Table 3. The specifics of the various exercises are presented later in this chapter.

Of the six exercises included in Table 3, the Conceptual Integration/Application, Tool and Object Nomenclature, Use, and Recognition, Dual Tasks, and Coordinative Speed and Accuracy exercises are based on the job learning concept. The Inspection/Sort exercise does not include a learning phase.

Table 3

Structure-of-Intellect Analysis of the Exercises Used to Assess the Learning and the Psychophysical and Motor Dimensions

	Psychophysical and Motor			*Learning*		
Contents / Operations	*Figural*	*Symbolic*	*Semantic*	*Figural*	*Symbolic*	*Semantic*
Cognition	2,5	2	-	4,6	4,6	1,6
Memory	5	-	-	6	1,6	1,6
Evaluation	5	-	-	3,4	1,3,4	1

Key to tests

1= Conceptual Integration/Application (troubleshooting)
2= Coordinative Speed and Accuracy (wiring)
3= Dual Task (control watch)
4= Dual Task (pipe assembly)
5= Inspection/Sort
6= Tool Object and Nomenclature, Use, and Recognition

Additionally, several exercises were designed to measure the social/motivational dimension. The performance tests developed to sample this dimension involve an actual group task (the Social Interactive Evaluation), a test of honesty (the Reliability Evaluation), and a game situation in which some intrinsic aspects of motivation are believed to be evidenced (the Level of Aspiration Evaluation). Finally, a semistructured interview was developed. The interview was designed to approach topics in three areas: (a) interest and background (e.g., areas, tasks, and jobs the individual likes to deal in or is drawn to), (b) motivation (e.g., does the individual being assessed possess goals and is he driven to attain these goals, ability of the individual being assessed to reconcile personal needs with those of the Navy), and (c) general capacity (e.g., common sense, alertness level, fund of general knowledge). Additionally, topics indicated by performance on the various

evaluative situations during the day were broached in the interview. The desired end product from the interview was an evaluation of each individual on each of several aspects (such as those given above) and a general evaluation.

Administrative Methods

In order to assure adequate professionalism and required standardization in the conduct of the TCAC, a formal Administrator's Manual (Wiesen & Siegel, 1975) was developed. The manual which was 116 pages in length contained details relative to:

a. The nature, scope, and purpose of the present program and background information on assessment centers.

b. Detailed administrative instructions for each exercise including, but not limited to: a precise, explicit script to be followed in the learning phase (where applicable), test administrative instructions, and timing and scoring instructions.

c. Samples of the various forms to be completed by the assessors relative to various exercises and judgments.

d. The schedule to be followed in the TCAC.

e. Privacy rights of the individuals assessed.

f. Methods of managing any special problems which might arise.

In addition to this Manual, formal assessor training was implemented. This training involved 3.0 days and included: (a) a detailed, formal review and elaboration of the materials in the manual, (b) practice exercise administration by each administrator with critique by exercise developers, (c) a dry run of one day's operation of the TCAC assessing Navy enlisted personnel as assessees and employing the precise methods to be employed in the conduct of the TCAC, and (d) review of the dry run by total TCAC staff.

Exercises

A summary of the formal exercises included in the TCAC is given in the subsequent paragraphs. Note that some of the exercises can be group administered, while others involve individual administration.

Conceptual Integration/Application (Troubleshooting) Exercise

- Purpose

The Conceptual Integration/Application evaluative situation was designed to measure capability to integrate simple facts and relationships with known information and to derive appropriate conclusions. To perform the task, a person must first be able to learn and remember certain rules. He must be able to perceive relationships and he must be able to manipulate several facts, rules, and relationships in his mind at the same time. This ability is often called "common sense." Within this evaluative situation, the operation/function/logic of a simple simulated electromechanical-hydraulic system is taught. Then malfunctions are inserted into the system. The task of the person being evaluated is to state the cause of the various malfunctions as they are presented. Guilford's Structure-of-Intellect model suggests that the following fundamental aspects of intellect are reflected by this task: cognition and memory operations, and symbolic and semantic contents.

- Outline of Procedure

A pretest instructional session provides all the needed information about a hypothetical system including the function of each part, the logic of operation, and the interrelationship among the parts. Various possible malfunctions and their probable cause(s) are also presented and briefly discussed. The test, which is oral in nature, consists of presentation of various system malfunctions for diagnosis by the individual(s) being assessed. The most likely cause(s) for each malfunction situation is indicated by the tested individual(s) on a formalized response sheet.

This evaluative exercise may be administered to groups of approximately 12 people. A total of about 50 minutes is required for the explanatory session and the formal test. Figure 2 shows this exercise being administered to three individuals.

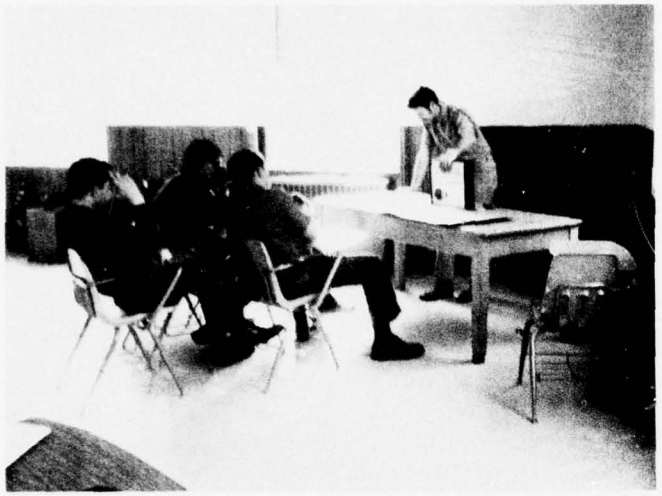


Figure 2. Conceptual Integration/ Application Administration.

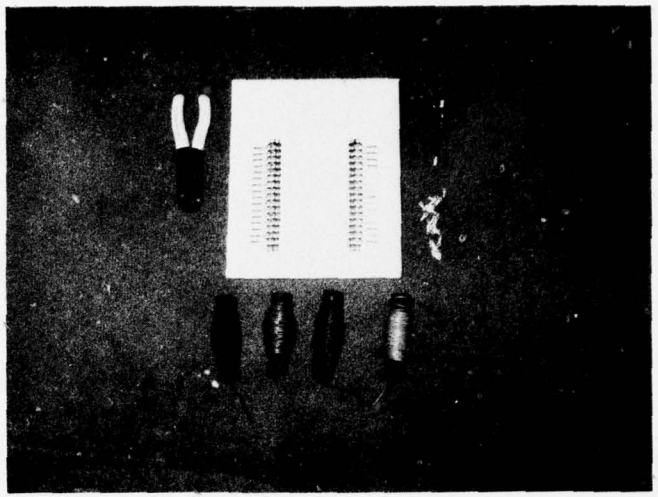


Figure 3. Tools and equipment for Co-ordinative Speed and Accuracy.

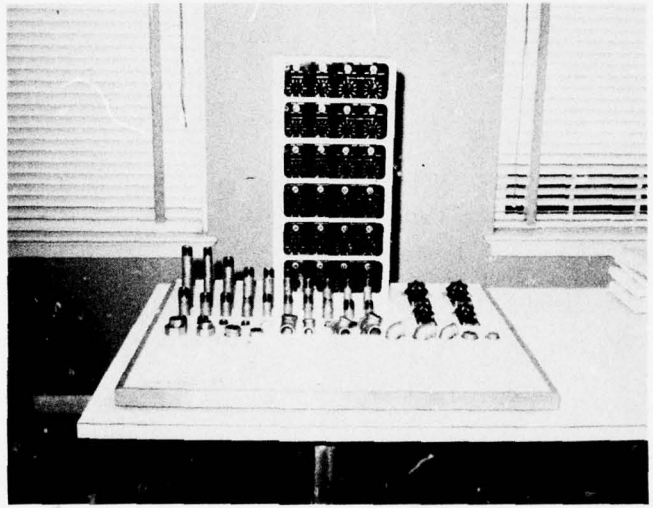


Figure 4. Dual Task equipment.

- Scoring Summary

The total number of individual and multiple correct responses is determined.

Coordinative Speed and Accuracy Exercise

- Purpose

The Coordinative Speed and Accuracy exercise was designed to assess ability to: organize and implement a systematic approach to work, pay attention to details, maintain short term and directed concentration on a task, and perceive relationships among parts to produce a required whole. Completion of the task requires working to achieve a goal having been given the parts. Accordingly, the situation allows for observations relative to an individual's approach to work.

This evaluative situation also involves the ability to use small tools, to make simple measurements, to work with small parts, and to perform precise work relative to a set of standards. These abilities are especially relevant to those Navy ratings in which independent work is involved and to ratings which require the use of small hand tools and manipulation of small parts.

The Guilford Structure-of-Intellect factors subsumed by this test are: cognition of figural/symbolic units, and cognition of figural/symbolic relations.

- Outline of Procedure

An instruction and practice session provides the individual with an opportunity to learn to perform a simple terminal board wiring task.

In the test situation, the individual is asked to put this new knowledge to use by connecting terminals on one terminal strip to terminals on a second terminal strip. These connections are made according to a wiring diagram and a color-coding chart using crimp-on connectors attached to 7" wire lengths. Figure 3 shows the tools and equipment supplied.

This evaluative situation may be administered to a group of up to four people. A total of about 35 minutes is required for the instruction, practice, and test periods. The test is administered as a timed test.

- Scoring Summary

The completed work is scored considering adequacy of the following: wire length, wire stripping, crimp-on connections, terminal screw tightening, choice of color of wire, adequacy of connections between terminals, and procedure employed in doing the task.

Separate scores are derived for speed (number of items completed), accuracy (percentage correctly completed), and a combination of speed and accuracy.

Dual Task Exercise

- Purpose

The Dual Task evaluative situation was designed to measure ability to perform in a time-sharing context. This type of task is common to a number of Navy watch standing situations. Performance of two unrelated tasks is involved in the Dual Task exercise. The situation requires the individual to keep track of and to perform two tasks concurrently. In the field of information processing, this is called time sharing. Time sharing is a part of Navy and of everyday life. The successful and joint completion of the tasks involved is dependent on this ability.

The two tasks involve simulated watch standing on a control-display panel and fabricating a pipe assembly. These particular tasks were chosen for two reasons. First, performance of tasks requiring the abilities subsumed by these tasks is likely to be required of many E4 level personnel. Second, the Structure-of-Intellect model developed by Guilford identifies several components of these tasks as being fundamental aspects of intellective performance. Using Guilford's terminology, these include the cognitive and evaluative operations as well as figural and symbolic content.

- Outline of Procedure

Simultaneously, and with a time limit, a pipe assembly is fabricated and a display-control panel is monitored. Several times, and in response to an indicator on the panel, the individual under assessment is required to make adjustments to the controls. The required control adjustments are made in response to a randomly appearing "warning." When the warning signal appears, the individual responds by making a set of precharted responses by serially "activating" eight control knobs. For the monitoring task, the amount of time taken to respond is scored along with the accuracy of control adjustments. The concomitantly performed pipe assembly task requires the individual being assessed to: (a) read a schematic, (b) measure lengths of pipe, (c) select the type of fittings and the diameter pipes and fittings indicated on the schematic, and (d) perform the actual assembly. Figure 4 presents the exercise equipment. The vertical panel shown in Figure 4 contains the control-display assembly, while the horizontal panel contains the necessary components for the pipe assembly task. A number of "distractor" (wrong) components is also included.

This evaluation is administered individually with a 12 minute time limit.

The pretest training situation for this exercise consists of a formal training period which involves instructions in both how to set the control panel when a warning signal occurs and how to read a schematic and make a pipe assembly. Hands on, one-to-one training and practice in pipe assembly is also given.

- Scoring Summary

For the monitoring task, the number of correct control settings and the integrated time to respond are scored. For the pipe assembly task, the number of correctly selected parts and the number of complete and correct connections are scored.

Inspectional/Sort Exercise

Purpose

The ability to make reasonable discriminations among meaningful stimuli is basic to all forms of learning and behavior. The Inspection/Sort evaluation assesses the ability to discriminate among objects that are similar to many found in Navy jobs. In addition to the ability to evaluate similarity, the ability to notice imperfections is evaluated.

This test requires the use of the cognition, memory, and evaluation operations coupled with figural content in the Guilford scheme.

Outline of Procedure

Ninety objects, 15 each of 6 types, are sorted by the individual into types (see Figure 5). A time limit is imposed. Some of the objects have imperfections or do not closely match the others. The individual being assessed places these into a reject bin. A "standard" comparison object of each type is displayed to the person being assessed at all times.

This evaluation may be administered to groups of any number of persons. The evaluation has a two-minute time limit.

Scoring Summary

Four scores are derived:

1. A speed score--number of items sorted/rejected.

2. An accuracy score for good items--percentage correctly sorted items (based on total items sorted).

3. An accuracy score for imperfect and mismatched items--percentage of correctly rejected items (based on items sorted).

4. A speed and accuracy score--number of items correctly sorted plus number of items correctly rejected.

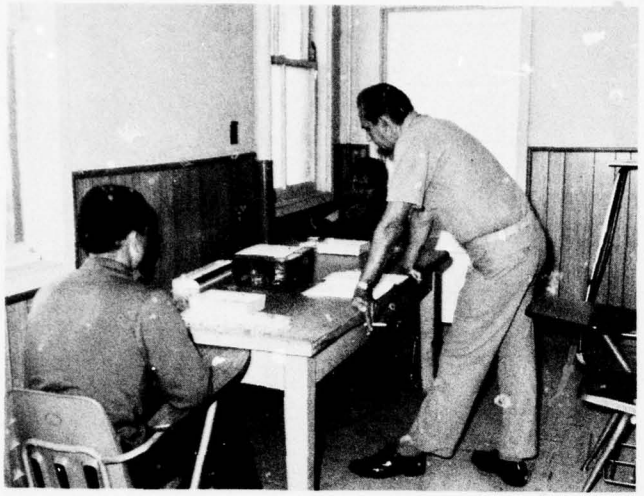


Figure 5. Inspection/Sort exercise administration.

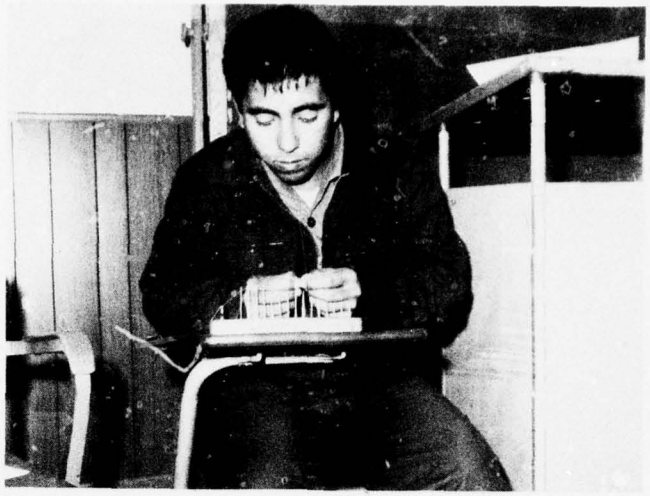


Figure 6. Reliability exercise

Figure 7. Tool and Object Nomenclature, Use, and Recognition administration.

Figure 8. Post exercise debriefing.

Level of Aspiration Exercise

- Purpose

The Level of Aspiration evaluative situation was designed to assess, among other things, need for achievement (the level of the goals that a person sets for himself). This, in turn, is reflected by general motivation and specific goal striving. Level of aspiration is generally more realistic in the more secure person. The reality orientation of the level of aspiration along with reaction to success and failure (e. g., tendency to give up easily) are also evaluated. The task also provides an opportunity for some informal conversation between the individual being assessed and the administrator. When this occurs, the ability of the individual to communicate, as well as his general demeanor, can be judged.

- Outline of Procedure

Behavior during a standardized dart throwing game provides the situational basis for the Level of Aspiration evaluation. A specific game sequence is followed which yields estimated and actual performance on the task. Specifically, the individual being assessed is asked: to make an initial performance estimate in view of a stated group performance average, to throw three darts and obtain an initial score, to make a second performance estimate in view of his actual performance, his original estimate, and the group average, and to throw three more darts; this procedure is repeated for three dart throwing trials. The group average, the estimates, and the attained scores are written on a blackboard which is visible to the individual being assessed, as the scores are obtained. These are permanently recorded at the conclusion of the test. Evaluation of behavioral traits is made at the end of the test using prepared scales and free form description (when appropriate).

This evaluation is individually administered and requires approximately 10 minutes.

- Scoring Summary

Each estimate and attained score is copied from the blackboard to a data sheet at the end of the evaluation. The initial estimates and the attained scores yield estimates of realism. Comparison of the given group average with the estimates of the individual being assessed yields a measure of need for achievement. Similarly, comparison of the attained and the predicted scores yields an estimate of realism.

Reliability Exercise

- Purpose

Individual morality and integrity are basic to the successful operation of our society including, of course, the Navy. Among psychologists, there is debate as to whether or not verbal self reports relative to morality will be reflected in actual behavior. In the present evaluative situation, this area of debate is bypassed by looking at a behavioral sample of actual moral behavior. The Reliability evaluative task is designed to provide a situation during which cheating is seemingly possible, undetectable, and beneficial. The individual is left alone during this evaluation. Accordingly, the moral decision to cheat or not to cheat is solely his, independent of the behavior, encouragement, or discouragement of others. Thus, this evaluative situation reflects intrinsic rather than extrinsic moral motivation.

- Outline of Procedure

In this evaluative situation (Figure 6), the individual is asked to perform an easy task (needle threading) which requires little instruction, to score his own performance, and to reset the apparatus to its original condition. While he is performing the task, the administrator attends to a visible and ostensibly pressing task. The evaluative task is graded in difficulty as a result of using a set of needles with graded eye size. The 10 needles with the largest eye size are rather easy to thread. The five needles with the smallest eyes are impossible to thread because their eyes are blocked with a transparent plastic. Therefore, the actual maximum score possible is less than the apparent maximum. If greater than 10, the reported score is fallacious.

This evaluative situation is individually administered with a 2.5 minute time limit. Figure 6 shows one person taking this exercise.

- Scoring Summary

The individual's self reported score is considered truthful if it is less than or equal to 10. A fallacious score can range from 11 to 15 depending upon how much the individual being assessed inflates his performance report over the attainable maximum of 10.

Social Interactive Exercise

- Purpose

When working in a group, some people make contributions that are more important and facilitating to the group goals than those made by others. Their ideas and suggestions are more likely to be accepted by the group than those made by other group members. People who have ideas and suggestions which are accepted by others are said to have leadership abilities.

The Social Interactive evaluation situation provides the assessor an opportunity to observe various individuals being assessed interacting in a standardized group task setting and to note the presence of leadership and positive social interactive skills in the individuals under assessment. The task is designed to provide an opportunity to see how well each individual cooperates and works within the group while the group is under goal achievement pressure.

The exercise situation is designed to involve increasing outside pressure for cooperation and leadership over trials. These behaviors are monitored by the administrator(s) and rated during each phase of the assigned group task.

- Outline of Procedure

The individuals under assessment are asked to perform a simulated resupply task, once with little urging and no time to contemplate cooperation, and twice with the strong suggestion and the time to plan a closely coordinated team effort. Data reflecting cooperation and leadership evidenced by each individual during each stage are recorded using a standard behavioral scoring system.

This exercise is administered to groups of three to six. The task requires five consecutive, three-minute periods.

- Scoring Summary

Observational data are recorded for each individual for each phase of the evaluation. Attention is given to behavioral indications of: leadership, cooperation, deportment, enthusiasm, ability to communicate clearly, attention to task, rule following, argumentativeness, and number and quality of ideas. Integrated scores and summaries of other comments are derived.

Tool and Object Nomenclature, Use, and Recognition Exercise

- Purpose

The Tool and Object Nomenclature, Use, and Recognition exercise is designed to assess several aspects of the ability to form and remember simple associations. The specific task chosen to allow these evaluations asks the individual to learn to associate the names of several objects taken from Navy life with the actual objects and their uses. Three recall situational aspects are measured: (a) recall of use based on observation, (b) recall of names based on observation, and (c) recall of names based on verbal description. These types of abilities are required to some extent in all Navy jobs. Guilford's Structure-of-Intellect model identifies several aspects of this evaluative task as fundamental to intellective function. Using Guilford's terminology, these include the operations: cognition and memory, and the contents: figural, symbolic, and semantic.

- Outline of Procedure

In a pretest learning situation, typical tools and objects taken from Navy life are presented. Their names and usage are described and briefly discussed. This presentation is followed by a verbally administered true-false test. Three scores are derived which reflect learning and memory of the information given. This evaluation may be administered to groups of up to about 12. A total time of about one hour is required to administer the two parts of this evaluation. Figure 7 presents the administration of a part of this exercise.

- Scoring Summary

Each of the three different types of questions is scored using its own key. The number of correct answers minus one-half the number of incorrect answers is calculated for each subtest. A grand total is also calculated.

Writing Sample Exercise (Optional)

- Purpose

Some ratings require a high level of writing ability, while others require only minimal writing skill. Only those individuals who express an interest in a rating which requires moderate to highly developed writing skills were asked to complete the Writing Sample exercise.

In the Writing Sample situation, the individual is given the opportunity to write a short (up to 300 words) response to two questions with a 30 minute time limit for both questions. One question is designed so that all individuals will have about equal knowledge on the subject. This question concerns boot camp activity. The second is designed to allow the individual to make some evaluative judgments concerning his present Navy assignment. Accordingly, in addition to the sample of writing behavior (which will provide measures of writing skill), the sample reflects some of the values and goals of the individual.

- Outline of Procedure

The individual under assessment is read and given a written copy of two questions. The individual has 30 minutes to write responses to these questions.

This evaluation may be administered individually or to groups.

- Scoring Summary

For each question, a score is obtained for: number of correctly spelled words, percentage of words correctly spelled, number of complete sentences, percentage of complete sentences, intelligibility, and insight.

Semistructured Interview

- Purpose

The purpose of the semistructured interview is to provide evaluative and predictive information not provided elsewhere in the information substrate for an assessee. The interviewer has available for his use all evaluative information generated up to the time of the interview of a given assessee. Additionally, the folder for a given assessee contains background data as available from the "service jacket" of the individual and related "basic battery" test results. The end result of the interview is a synthesis of these materials into a meaningful nexus with additional information provided which will fill in gaps relative to the purposes of the Center--providing, if possible, a "career field" recommendation for a given assessee.

The interviewer is a person who is trained in and skilled at interview methods. Additionally, he has experience and knowledge in personality dynamics, motivational theory, and the like.

- Content of Interview

Because it is important that all interviewers use somewhat the same approach and questions, a semistructured format is involved. This helps to assure between interviewer reliability in content. However, any area of import vis-a-vis a specific assessee is pursued in sufficient depth to satisfy the interview purposes. The three general areas which are probed in each interview are: interests/background, motivation, and general capacity.

- Interview Report

An interview checklist rating form is completed. Each of 16 items is rated on a 4 point scale. A checkoff is provided for items that cannot be rated based on the interview. Space is provided at the end of the checklist form to enter any justifications (examples) for the ratings and, more importantly, for entering any additional information which may be important for the use of the panel when it considers the individual under evaluation.

Briefing/Debriefing and TCAC Evaluation Form

Two group meetings, one at the beginning and one at the end of each day on which the TCAC operated, were held to answer any questions of the individuals under assessment and to review the nature, scope, and purpose of the program.

At the end of the day's meeting (Figure 8) each group of individuals under assessment was asked to provide verbal and written feedback about the several aspects of the Center's operations, including: perceived fairness of the tests and relative preference for the various tests. A standard response form was developed for this purpose. The morning meeting included a discussion and a verbatim reading of a statement concerning rights to freedom of information and privacy.

Administrative Procedures for TCAC

Wiesen and Siegel (1975) described, in detail, all aspects of the operation of the TCAC, including, but not limited to, preparing for, administering and scoring the exercises, scheduling, and TCAC decision procedures. Briefly stated, the daily procedure involved: briefing the group of individuals to be assessed that day relative to the purpose, content, and goals of the TCAC (15 min.), administration of the set of exercises (5 hours), and debriefing (15 min.).

Weekly meetings of the Center's staff were held during which each person under assessment was individually considered and in which a classification and a set of ratings were completed. The procedure for this weekly meeting was prescribed in advance in the administrator's manual (Wiesen & Siegel, 1975). To facilitate these weekly discussions, there was a meeting of the Center's assessors at the end of each day. In this meeting, each individual tested that day was assigned to one administrator who would serve as an advocate for that individual at the subsequent weekly meeting. The advocate for an individual summarized the relevant data, gave his evaluation of the individual on the three dimensions discussed earlier, and made an initial proposal of career field for that person. The discussion which followed always began with a statement from the assessor who interviewed the individual being assessed. Then, other contributions of information or observations concerning the individual were made by the other assessors. A discussion of the merits proposed for the individual then ensued. This

was followed by a secret ballot. Each administrator voted either to accept or to reject the proposal and to make a prediction of the anticipated level of on-the-job success in that rating. In the event of a defeated proposal, one of two avenues could be followed: (a) a second proposal could be made, or (b) if there was no second proposal, the individual was recommended for general rates. In the event of a tie vote, a second discussion period and vote occurred. If there was still a tie vote, then the Center's director cast a "tiebreaker" ballot.

Individuals Assessed

Enlisted personnel (N = 140) in the Fireman or Seaman Apprentice Training Schools at the Naval Training Center, San Diego, California, were assessed by the TCAC. Within limits of a practical nature, the individuals were chosen to be not A school qualified four year obligators, with basic battery GCT and ARI scores totaling less than 100.

Description of Sample

Mean sample scores on the General Classification Test (GCT), Arithmetic (ARI), Mechanical (MECH), Clerical (CLER), Electronic (ETST), and Shop (SHOP PR) tests are given in Table 4. These tests were administered by the Navy to all enlisted personnel on entry into the Navy and are designed to have a mean of 50 and a standard deviation of 10. The means for the present sample are less than 50 for all but the SHOP PR test. Four of these means (GCT, ARI, MECH, and ETST) are significantly below 50. The group standard deviations are less than 10 for all six tests. These differences from total norms were to be anticipated. The population of enlisted men in the Apprentice School is generally restricted in range to the lower half of the distributions of these "basic battery" test scores. The higher scoring men typically are assigned to an A school. Thus, the desired sample of low aptitude enlisted personnel was assessed.

The group received a mean of 2.1 demerits in basic training. This demerit level is not considered to be unusual.

The mean age, level of schooling, and reported arrest record were, respectively, 19, 12, and 21. Accordingly, the sample contained, on the average, recent high school graduates, about 21 per cent of whom reported an arrest record.

Table 4

Mean and Standard Deviation of Sample on Navy Related Scores and Biographical Items

Variable	*Mean*	*Standard Deviation*	*Notes*
NAVY RELATED SCORES			
1. GCT	47.64	5.12	mean differs from 50, t (139)= -5.43, $p < .01$
2. ARI	46.27	5.85	mean differs from 50, t (139)= -7.56, $p < .01$
3. MECH	48.01	6.74	mean differs from 50, t (139)= -3.48, $p < .01$
4. CLER	49.43	6.44	mean does not differ from 50, t (139)= -1.04, $p > .05$
5. ETST	47.61	9.96	mean differs from 50, z= -2.83, $p < .01$
6. SHOP PR	51.09	7.33	mean does not differ from 50, t (139)= 1.75, $p > .05$
7. Demerits in first weeks of training	2.14	3.46	
BIOGRAPHICAL ITEMS			
8. Schooling	11.87	.97	highest school grade completed
9. Arrest record	.21	.41	scored dicotomously: arrest= 1, no arrest= 0
10. Age	19.03	1.65	age to nearest birthday

TCAC Staff and Time of Operation

The Director of the TCAC and one other administrator were psychologists. The remaining two administrators were Navy Chief Petty Officers. It was believed that this administrator mix would provide a balanced staff of persons who represented both the Navy job requirements and the personality and individual traits points of view. The Director's position was filled by two people during the course of operation of the Center to make a total of five administrators who were directly associated with the day-to-day operations. The TCAC was in operation for approximately seven weeks during the months of November and December, 1975.

Data Considered

The data considered for each individual assessed and classified included 28 scores resulting from the Center's exercises, biographical data, observations, 7 scores taken from Navy records, and 10 scores reflecting the Center's overall evaluation. The substance of these data is summarized in Table 5.

Table 5

Number and Source of Data Scores Recorded for Each Individual Assessed/Classified

Source	Number of Scores	Source of Scores
Center Exercise	1	Conceptual Integration/Application number of correct troubleshooting items
Center Exercise	3	Coordinative Speed and Accuracy speed accuracy speed and accuracy
Center Exercise	3	Dual Task pipe assembly (total) control watch (number of correct settings and latency)
Center Exercise	4	Inspection/Sort speed score accuracy score for good times accuracy score for imperfect/mismatched items speed and accuracy score
Center Exercise	7	Level of Aspiration first estimate by individual sum of remaining estimates sum of realism scores sum of pessimism scores realism evaluation (dichotomous) optimism evaluation (dichotomous) striving evaluation (dichotomous)
Center Exercise	1	Reliability lie-no lie dichotomy
Center Exercise	3	Social Interactive phase 1-3 total phase 4-9 total total
Center Exercise	4	Tool and Object Nomenclature, Use, and Recognition observation recall--use observation recall--name verbal recall--name total
Center Exercise	-	Writing Sample This exercise was optional. Only 13 assessees elected this exercise. It is not considered further.
Center Observation	2	Interview weighted sum of evaluations number of items not judged
Center Observation	-	Incidental Observations
Center Observation	-	Biographical Data educational background arrest record
Navy Records	7	Data from Navy Records test scores (basic battery) basic training disciplinary record
Center Evaluation	1	career field proposed (yes-no dichotomy)
	1	career field proposed (specific field)
	4	forecast of on-the-job performance level: 4 point scale (recorded separately for each administrator)
	1	number of proposals considered for an individual
	3	evaluation of each assessee on each of three dimensions: learning, psychomotor, and social

RESULTS AND INTERPRETATION

Career Field Recommendations

The TCAC arrived at two decisions relative to each man processed: a career field recommendation, and a prediction of job success in that career field after six months on the job.

Career field recommendations were based on the evaluations and on the career field requirements as determined by the job analysis described in the introductory chapter.

The career field recommendations of the TCAC staff were quite diversified given the background of the men in the sample. All of the enlisted men assessed had been previously selected for general duty on completion of the Apprentice School. These men were considered on the basis of the usual classification procedure to be generally unqualified for immediate entry to a specific rating through A school training. Of the total sample assessed, 79 per cent had achieved a combined GCT and ARI score of less than 100. Despite this, the TCAC recommended 54 per cent of the men assessed for career fields (specific ratings). This suggests some sensitivity to the TCAC process not found in the usual Navy classification process. The remaining 46 per cent were recommended to continue in the general rates. The specialty most often (27 per cent) recommended was Machinist's Mate. The complete distribution of recommendations is given in Table 6.

Predictions of On-the-Job Success

Each administrator made a prediction of on-the-job success for each individual assessed relative to the assigned career field. The administrators made this rating on the basis of their interpretation of the data available for each individual assessed and classified. These predictions, made on a scale ranging from 0 to 3, were summed to get an overall level of on-the-job success prediction for each assessee. This combined score could range from 0 to 12.

Preceding page blank

Table 6

Career Field Recommendations and Corresponding Mean Predictions of Job Success

Career Field	*Number Recommended*	*Percent Recommended*	*Mean Predicted Level of on the job Success**
General Rates	64	45.7	8.6
Postal Clerk (PC)	11	7.9	10.2
Storekeeper (SK)	14	10.0	10.9
Yeoman (YN)	6	4.3	8.5
Signalman (SM)	2	1.4	9.0
Machinist's Mate (MM)	38	27.1	9.8
Hospital Corpsman (HM)	2	2.1	9.5
Other Specific Rating	3	1.4	11.7
TOTAL	140		weighted mean 9.4

**scores range from 0 to 12*

The mean level of success for those men recommended for general rates and for the men recommended for specific ratings are, respectively, 8.6 and 10.1. Details of these predictions are given in Table 6. The mean prediction of success for the men recommended for general rates was neither the lowest nor the highest of the mean predictions for the various career fields. It was, however, lower than the means of most of the groups of men recommended for a specific career field. This lack of a simple relationship between the level of difficulty of the job and the predicted level of on-the-job performance appears to be due to joint consideration of the man and the career field recommendation made for that man. The specific career field was chosen based on the man's global characteristics. The prediction of level of job success was then made for the man in the rating suggested. For example, a man who was evaluated highly by the TCAC and therefore recommended for a rating (e.g., HM) might be predicted to show an average level of on-the-job performance due to the demanding nature of the rating recommended. Due to this, the subsequent analysis of predicted level of job performance includes consideration of men within ratings rather than simply considering together all men in all ratings.

Exercise Scores and Dimensional Ratings

The mean and standard deviation of each set of TCAC exercise scores/judgments was calculated. A description of these scores, along with the obtained means and standard deviations, is presented in Table 7. We note that there are no indications of the means being affected by either a ceiling or a floor effect. Moreover, in each case, the standard deviation seems reasonable. Quite obviously, there is no normative group with which these distributions may be compared.

The product moment correlations along all scores were also calculated. These intercorrelations were generally moderate but not high. This indicates some uniqueness for the various measures, as well as a degree of overlap. Subscores within some tests seem to be highly intercorrelated. Accordingly, a subset of 16 scores was chosen to be used in most of the further analyses. These 16 scores are marked with an asterisk in Table 7.

The intercorrelational matrix that includes these 16 TCAC measures, as well as the Navy related scores, biographical items, and the several TCAC judgments is presented in Table 8.

Dimensional Judgments

Each person processed was rated on each of the three dimensions--learning, psychophysical and motor, and social/motivational--employed as a basis for the job analysis. The judgments relative to each individual assessed were made by only one administrator. The mean across assesses, for each of the three dimensions, was, respectively, 3.18, 3.23, and 3.19. These means are close to the midpoint (3.00) of the 1 to 5 rating scale used and are quite close to each other. This suggests that there was no overall bias on the part of the administrators to rate one dimension more highly than another.

Table 7

Mean and Standard Deviation (S. D.) of TCAC Judgment and Exercise Scores

Judgments	*Mean*	*S.D.*	*Comment*
Forecast	9.87	2.36	prediction of on the job success in recommended career field
Number of Discussions	1.09	.38	the number of discussions and ballot sessions necessary for a classification decision
Administrator Comment Forms	-.01	.33	positive comments= +1, negative comments= -1
TCAC Rating: learning dimension	3.16	1.03	
TCAC Rating: psychophysical and motor dimensions	3.25	.88	
TCAC Rating: social/ motivational dimension	3.19	.84	
EXERCISE SCORES			
*Conceptual Integration/ Application	7.96	6.03	troubleshooting task
Inspection/Sort: score 8	43.40	16.31	speed score
*Inspection/Sort: score 9	77.33	18.75	accuracy score for imperfect/mismatched items
*Inspection/Sort: score 10	98.11	3.22	accuracy score for good items
*Inspection/Sort: score 11	40.48	14.95	combined speed and accuracy score
*Reliability	.09	.29	
Tool and Object: score 1	19.12	1.73	recall of uses based on observation
Tool and Object: score 2	13.11	2.64	recall of names based on observation
Tool and Object: score 3	12.20	2.40	recall of names based on verbal description
*Tool and Object: total score	38.42	5.05	
*Dual Task: score 1	22.56	6.45	construction score
*Dual Task: score 2	38.86	5.35	accuracy of control settings
*Dual Task: score 3	248.28	195.22	latency to respond
Coordinative Speed and Accuracy: score 1	47.77	11.71	wiring: combined speed and accuracy score
*Coordinative Speed and Accuracy: score 2	51.31	10.58	wiring: speed score
*Coordinative Speed and Accuracy: score 3	92.56	8.80	wiring: accuracy score
*Level of Aspiration: score 1	67.04	19.21	level of aspiration - initial
Level of Aspiration: score 2	216.07	59.67	level of aspiration - after experience
*Level of Aspiration: score 3	19.81	47.34	realism score
Level of Aspiration: score 4	1.25	1.06	pessimism score
Level of Aspiration: score 5	.89	.31	realism rating
Level of Aspiration: score 6	.71	.45	optimism rating
Level of Aspiration: score 7	.94	.25	striving rating
Group Task: score 1	2.38	4.42	first phase score
Group Task: score 2	5.44	4.47	second phase score
*Group Task: score 3	7.80	6.31	overall score
*Interview: score 1	2.62	.37	mean of items rated (on a scale from 1 to 4)
*Interview: score 2	1.74	2.11	number of items not rated

**TCAC test Scores chosen for further analysis.*

Table 8

Intercorrelations Among Various Measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1		03	18	08	12	20	-04	-21	18	-20	-05	01	23	01	07	13	00	-01	-05	08	25	15	15	-21	11	02	-09	09	00	14	-09
2			08	12	30	07	-09	-15	05	-11	14	02	04	15	05	01	-07	19	-01	-07	07	22	03	04	10	21	-03	-12	01	12	16
3				-16	14	35	-01	-05	-13	-01	12	14	40	38	-01	28	22	11	03	-04	45	31	09	-07	17	19	08	-10	03	07	-06
4					04	-01	-11	-07	10	04	22	04	02	08	08	09	01	04	-04	-01	-03	13	10	-09	10	05	-02	00	-01	05	13
5						26	00	15	03	03	07	-05	19	11	-04	09	08	07	-01	08	14	13	06	16	05	02	01	-16	-04	06	-01
6							-04	05	-04	-01	26	16	36	29	06	23	11	10	04	04	42	30	06	-08	18	13	01	01	14	10	-12
7								-12	-02	-17	-07	-13	-08	-18	02	-09	-12	-17	-21	02	-02	-25	01	-03	-19	-20	-06	-02	04	-26	06
8									-09	55	00	03	-12	-13	-06	-13	06	-06	12	-06	-06	-13	-12	15	-17	-13	-05	-04	-06	17	-06
9										02	-06	05	08	-01	15	04	-05	12	-04	-04	03	05	14	-18	15	03	01	08	13	-07	04
10											-05	00	-14	-11	05	-15	02	07	09	-05	-14	-08	-20	15	-02	-04	05	12	02	13	06
11												29	38	31	35	23	03	14	11	-25	23	35	14	-05	12	32	-05	-26	19	18	-12
12													46	29	45	47	-06	-05	16	-33	29	11	13	-18	27	29	12	-08	24	27	-14
13														42	27	71	15	12	12	-08	60	41	18	-17	30	33	09	-13	11	15	-25
14															25	23	02	23	40	-20	29	61	16	-15	56	39	18	-12	14	03	-08
15																20	-12	-01	20	-26	11	16	18	-20	14	27	11	00	51	28	-04
16																	11	07	-02	01	43	14	08	-11	20	34	15	-02	09	11	-10
17																		-09	-24	11	02	10	00	-04	04	-05	05	07	01	05	-17
18																			27	00	03	17	07	-03	15	33	01	04	06	-01	08
19																				-08	03	23	-03	06	23	14	16	02	07	14	-01
20																					-05	-11	-01	14	-15	-02	00	16	-08	-05	11
21																						30	23	-23	32	30	06	-17	09	13	-08
22																							15	-21	45	32	14	-12	07	07	-21
23																								-71	22	26	04	-11	11	17	-01
24																									-23	-28	-13	-01	-15	-12	-01
25																										27	24	01	09	08	-24
26																											31	-03	05	14	04
26																												43	09	13	-16
28																													05	06	24
29																														09	-07
30																															-15
31																															

KEY

1= GCT, *2*= ARI, *3*= MECH, *4*= CLER, *5*= ETST, *6*= SHOP PR, *7*= Demerits in first weeks of training, *8*= Highest School Grade, *9*= Arrest record, *10*= Age, *11*= General rates - Specialty dichotomy, *12*= Prediction of the job success, *13*= TCAC Rating: learning dimension, *14*= TCAC Rating: psychophysical and motor dimension, *15*= TCAC Rating: social/motivational dimension, *16*= Conceptual Integration/Application, *17*= Inspection/Sort: accuracy , imperfect/mismatched items, *18*= Inspection/Sort: accuracy, good items, *19*= Inspection/Sort: speed and accuracy, *20*= Reliability, *21*= Tool and Object: total, *22*= Dual Task: construction score, *23*= Dual Task: control settings, *24*= Dual Task: latency, *25*= Coordinative Speed and Accuracy: speed, *26*= Coordinative Speed and Accuracy: accuracy, *27*= Level of Aspiration: initial estimation of goal level, *28*= Level of Aspiration: realism score, *29*= Group Task: total, *30*= Interview: mean rating, *31*= Interview: number of items not rated.

NOTE: Decimal points omitted.

Interrater Reliability

The interrater reliability of the administrators was investigated. Four administrators each scored the work of from 5 to 19 assessees. An intraclass correlation coefficient (the average of the correlation coefficients between pairs of administrators) was used to index the amount of agreement, by exercise, between the scores reported by each administrator for each individual assessed (Guilford, 1965, pp. 299-300).

The level of reliability could not be investigated in this way for all measures. Some of the measures for some of the exercises required dismantling or otherwise destroying the work of the individual being assessed. The approach to reliability used with these measures was to derive scores for those aspects of the exercise that were scorable without dismantling or destroying the product to be scored.

The intraclass correlation coefficients indexing interrater reliability ranged from .72 to 1.00 with a median of .97 and an interquartile range from .92 to .99. These reliability coefficients are presented in Table 9. The one aberrantly low reliability coefficient of .72 (accuracy score for dual task exercise) was based on the fewest number of subjects (five). In fact, one administrator's relatively large discrepancy regarding the score of one assessee caused this lowered reliability coefficient. All remaining 19 scores involved in this reliability coefficient were unanimously agreed on by the four administrators.

These levels of reliability compare quite favorably with those found in the literature. The scores were all of tangible end products. Working with tangible end products, Bornstein, Jensen, and Dunn (1954) reported a mean phi coefficient reflecting interrater reliability of .78 (with a standard deviation of .19).

Table 9

Interrater Reliability (Intraclass Correlation) of Selected Center Exercise Scores

Measure	*Number of Subjects Represented*	*Number of Administrators Represented*	*Reliability Coefficient*
Coordinative Speed and Accuracy: combined and accuracy score*	19	4	.96
speed score*	16	4	.97
accuracy score*	16	4	.89
Dual Task			
construction score	6	4	.99
accuracy of settings	5	4	.72
Inspection/Sort:			
speed score	15	4	1.00
accuracy score	15	4	.93
accuracy score	15	4	.92
combined speed and accuracy score	15	4	.99

**Certain measures in this test require dismantling the assessee's work. Only those measures that are nondestructive were scored for this reliability estimate.*

Dimensional Evaluations

The three dimensional ratings were further considered in three ways: (1) inter and intradimensional relationships of the measures, (2) consistency of the policies of the individual administrators, and (3) degree of agreement among the administrators. These two types of accord (intra-administrator consistency and interadministrator agreement) reflect different measurement constructs. Intra-administrator consistency is a measure of the consistency of any one administrator in applying his decision making policy. The interadministrator agreement is a measure of how closely the policies of the individual administrators agree. This administrator agreement may also be taken as an indication of the interadministrator reliability relative to these ratings. A policy capturing approach to decision making was employed in the analysis of these two types of accord.

Interdimensional Relationships

Three Pearson product moment correlation coefficients were calculated considering all possible pairs of dimensions. These intercorrelations are given in Table 10. The common variability in these dimensions, as indexed by r^2, ranges from 6 per cent to 18 per cent. The highest relationship was between the learning and the psychophysical and motor dimensions. The dimensional measurements seem, accordingly, to have been independently considered by the administrators.

Table 10

Intercorrelations Among Dimensional Ratings

	Learning	Psychophysical and Motor	Social/ Motivational
Learning	-	.42	.27
Psychophysical and Physical		-	.25
			-

Intradimensional Relationships

The various TCAC exercises were intended to measure abilities along three dimensions--learning, psychophysical and motor, and social/motivational. The exercises designed for the TCAC were mainly complex job sample learning and performance tests. They were not designed to measure singular or simple abilities. As indicated in Table 3, most of the exercises were designed to span Guilford Structure-of-Intellect categories. Accordingly, some intercorrelations between any two exercises might be expected with the highest intercorrelations occurring among exercises reflecting any one dimension. This, indeed, was the case.

A comparison was completed of the proportion of statistically significant product moment correlations among measures of one dimension with the number of statistically significant correlation coefficients between two dimensions. The critical value with 140 assesses for $p = .05$, two tailed is $r = .17$. Of the total of 378 possible comparisons (28 scores taken two at a time), 251 are interdimensional and 127 are intradimensional correlations. Of the interdimensional correlation coefficients, 20.3 per cent (51) were statistically significant. Of the intradimensional correlations, 40.2 per cent (51) were statistically significant. This general trend supports arguments favoring the fidelity of the exercises for measuring the dimensions involved. Each dimension is treated in more detail below.

The Pearson correlation coefficients among the scores designed to reflect specifically the learning dimension are given in Table 11. All but 7 of the 28 correlation coefficients of Table 11 are statistically significant. These statistically significant correlation coefficients range from .17 to .80, with an interquartile range from .22 to .43. These correlation coefficients indicate that the learning oriented tests are, by and large, moderately related. This finding was expected because the exercises range in form from a conceptual troubleshooting task to a timed inspection test, to a timesharing task including a pipe assembly and a control watch. The high percentage of statistically significant intercorrelations can be taken as an indication or confirmation of the ability of these exercises to measure various aspects of a common, albeit complex, dimension.

Table 11

Intercorrelations Among TCAC Scores Relevant to Learning Dimension

	Score	1	2	3	4	5	6	7	8
1	Conceptual Integration/Application	-	.28*	.33*	.39*	.43*	.14	.08	.11
2	Tool and Object: use--visual		-	.43*	.45*	.75*	.32*	.16	.17*
3	Tool and Object: name--visual			-	.37*	.78*	.18*	.22*	.16
4	Tool and Object: name--description				-	.80*	.22*	.16	.20*
5	Tool and Object: total					-	.30*	.23*	.23*
6	Dual Task: construction score						-	.15	-.21*
7	Dual Task: control setting							-	.71*
8	Dual Task: latency								-

* *p* < .05, *two tailed test*

NOTE: Signs of some correlation coefficients changed so that positive relationships indicate better performance on both scores.

Psychophysical and Motor Dimension

The Pearson product moment correlations among the scores designed to measure the psychophysical and motor dimension are given in Table 12. All but one of these scores was moderately related with the other scores. Some accuracy scores on the Inspection/Sort exercise were negatively correlated with scores that reflect speed in the Inspection/Sort exercise. This tradeoff between speed and accuracy is not unusual.

Of the 21 correlation coefficients between these scores, 14 (66.7 per cent) are significant. This seems considerably higher than the 20.3 per cent significant correlation coefficients in interdimensional correlations, as described in a previous section. The range of statistically significant correlation coefficients is from .20 to .98. The interquartile range is .23 to .57.

These moderate, positive correlations can be construed as support for the conjecture that these exercise scores reflect aspects of one underlying dimension.

Social/Motivational Dimension

The intercorrelations among the scores designed to consider the Social/Motivational dimension are given in Table 13. For the 16 correlation coefficients that differed significantly from zero, 11 are positive and 6 are negative. The negative correlations all reflect differences in scores on the Level of Aspiration exercise. The pessimism scores correlated negatively with some of the other scores (the first goal attainment level estimate and the optimism rating). There was also a negative correlation between the realism score and the rating of optimism. Perhaps these scores cannot vary independently in a task such as this one; that is, a highly optimistic estimate cannot be realistic or an individual cannot be optimistic and pessimistic at the same time.

The overall low level of interrelationship between some of the scores on the Level of Aspiration exercise and the other measures of this dimension reflects the uniqueness for these measurements. Of the 78 correlations within the Social/Motivational dimension, 16 (20.5 per cent) are statistically significant. This value may be compared with 20.3 per cent significant correlation coefficients among all the interdimensional correlations. This moderate level of agreement within these scores may be taken as support for the conjecture that these measures involve a common construct.

Table 12

Intercorrelations Among Scores Relevant to Psychophysical and Motor Dimension

	Score	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1	Inspection/Sort: speed score	-	-.34*	.20*	.98*	.24*	.22*	.12
2	Inspection/Sort: accuracy, imperfect items		-	-.09	-.24*	.01	.04	-.05
3	Inspection/Sort: accuracy, good items			-	.27*	.24*	.15	.33*
4	Inspection/Sort: speed and accuracy				-	.25*	.23*	.14
5	Coordinative Speed and Accuracy: speed and accuracy					-	.94*	.57*
6	Coordinative Speed and Accuracy: speed						-	.27*
7	Coordinative Speed and Accuracy: accuracy							-

*$p < .05$, two tailed test

Table 13

Intercorrelations Among Scores Relevant to Social/Motivational Dimension

	Score	2	3	4	5	6	7	8	9	10	11	12	13
1	Reliability	.00	-.03	-.16	-.08	-.13	.04	-.02	-.05	-.08	-.08	-.05	.11
2	Level of Aspiration: score 1-initial estimate of goal level	-	.39*	-.43*	-.31*	-.15	.38*	-.04	-.03	.12	.05	.14	.04
3	Level of Aspiration: score 2 - sum of subsequent estimates		-	-.15	-.06	-.07	.33*	.17*	-.01	.09	.09	.13	-.16
4	Level of Aspiration: score 3 - realism score			-	.65*	.29*	-.40*	.05	.05	-.09	-.05	-.06	-.24*
5	Level of Aspiration: score 4 - pessimism score				-	.08	-.48*	-.08	-.02	-.01	-.01	-.15	-.18*
6	Level of Aspiration: score 5 - realism rating					-	-.01	.00	.02	.03	.02	.15	-.12
7	Level of Aspiration: score 6 - optimism rating						-	.16	.07	.16	.14	.20*	-.02
8	Level of Aspiration: score 7 - striving rating							-	.01	.00	.00	.01	.09
9	Group Task: phase 1								-	.51*	.57*	.07	-.08
10	Group Task: phase 2									-	.96*	.11	-.03
11	Group Task: total										-	.09	-.07
12	Interview: mean rating											-	-.15
13	Interview: number of items not scored												-

Signs of some correlation coefficients changed so that positive relationships indicate increasing magnitude on the two scores.

$p < .05$, two tailed tests.

Policy Capturing of Dimensional Ratings

One of the intermediate steps in the classification decision making process was rating each man relative to each of the three dimensions considered by the job analysis. One administrator rated each man. The policy of the individual Center administrators and the overall policy of the administrators were investigated. The goal of these policy capturing analyses was to determine the extent to which the ratings of the Center can be seen to be themselves predicted by a weighted sum of individual test scores. Stated alternatively, the policy capturing approach seeks to establish the extent to which the tests captured the decision making policy of the administrators. To this end, a stepwise multiple linear regression analysis procedure was used. This procedure allows the evaluation of each score relative to the decision of the administrator(s).

The policy capturing approach to decision analysis possesses a fairly recent history of application. Madden (1964), Stephenson and Ward (1971). Bottenberg and Christal (1968), and Christal (1968a, 1968b, 1963) described its employment for Air Force officer advancement evaluative purposes, and Siegel and Federman (in press) employed the approach for deriving emphasis areas in Air Force Technical Training. The approach has also been applied in a wide variety of other areas including, but not limited to, judgments of personality characteristics (Hammond, Hursch, & Todd, 1964), attraction of common stocks (Slovic, 1969), mental illness diagnosis (Goldberg, 1970), and judgments of admissibility to graduate school (Dawes, 1970, 1971). While there is some theoretic controversy relative to the use of the additive model for such work, Slovic and Lichtenstein (1970) after a comprehensive review of studies employing the linear approach, concluded:

> In all of these situations the linear model has done a fairly good job of predicting the judgments, as indicated by r_s values in the .80s and .90s for the artificial tasks and the .70s for the more complex real-world situations (p.36).

If acceptable multiple correlation coefficients are evidenced, then it can be said that the various scores are actually reflective of the policy employed by the administrator(s) in making a rating.

The regression equation gives the weights seemingly used by the rater. The square of the multiple correlation coefficient gives the proportion of variance the policy accounts for. In the analyses here completed, only the original scores were used. No attempt was made to introduce nonlinear terms into the regression equation by generating additional scores from the raw scores. Of course, from a theoretical point of view, the ability to predict a decision using a mathematical formula does not mean that the actual decision making process has been identified. However, from a pragmatic point of view, the decision can be said to be understood and captured by such a formula.

Policy in Rating Men on Learning Dimension

Following Cristal (1963), the individual policies of the five administrators should all be more fully captured by a multiple regression approach than the one overall policy of the five administrators considered together. The policies of the individual administrators were indeed captured to a greater extent for the learning dimension than the group policy. However, for three of the administrators, the N was small relative to the number of tests. The multiple R for these administrators must be interpreted accordingly. Stepwise multiple regression analyses reflecting the policies of the five individual administrators were conducted to reveal the tests most emphasized by each administrator, relative to the learning dimension.

The multiple correlation coefficients for the five administrators were: 1.00, 1.00, .95, .91, and .85. These indicate a very considerable predictiveness of the rating by the exercise scores. Additionally, if the multiple correlation can be considered to represent valid covariance, then these values provide some measure of intrarater (rate-rerate) reliability. Moreover, this rate-rerate reliability measure is derived without the negative effects of an administrator rating the same person twice.

The extent to which there is agreement in policy relative to judgments on the learning dimension among the administrators is indicated by the extent to which the grouped data of these five administrators can be described by one policy.

The overall group policy concerning the learning dimension was evaluated through a stepwise multiple linear regression analysis. The group policy emphasized the Conceptual Integration/Application; Tool and Object Nomenclature, Use, and Recognition; and the Dual Task construction scores. The other scores did not contribute substantially to the explanation of the group rating policy on this dimension. The analysis is summarized in Table 14. The very high multiple correlation obtained for this dimension (R = .84) may be interpreted, according to the literature, as indicating high agreement among the policies of the individual members of the TCAC staff (Christal, 1963).

Table 14

Group Policy for Rating Men on the Learning Dimension

Step	*Score*	*Stepwise Multiple R*	R^2	*Increase in* R^2	*Zero Order r*
1	Conceptual Integration/Application	.71	.51	.51	.71
2	Tool-Object Nomenclature, Use, and Recognition	.78	.61	.10	.60
3	Dual Task: construction score	.82	.67	.05	.41
4	Interview: score 2	.83	.68	.01	-.25
5	Inspection/Sort: speed and accuracy score	.83	.69	.01	.12
6	Inspection/Sort: accuracy score	.83	.69	0*	.15

Multiple R based on Center test scores= .84
**less than .005*

The group policy is also summarized in Table 15 for ease of comparison with the policies of the individual administrators. Table 15 indicates that two of the three tests most emphasized by each individual policy coincide with two of the three tests most emphasized by the group policy.

Table 15

Relative (Order of Regression Entry) Importance of Center Scores Most Weighted in the Policies of Each Administrator and of the Administrators as a Group

TCAC SCORE	LEARNING DIMENSION						PSYCHOPHYSICAL-MOTOR DIMENSION						SOCIAL-MOTIVATIONAL DIMENSION					
	Group Policy	Administrators					Group Policy	Administrators					Group Policy	Administrators				
		1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
Conceptual Integration/Application	1	2	3	4	1	1				5	4		6	2		2	6	
Inspection/Sort: accuracy, imperfect	6	4	4		6				4	6			5					3
Inspection/Sort: accuracy, good								4								5		
Inspection/Sort: speed and accuracy	5			5	4	6	3	2	3	3	3					4	4	
Reliability							5	6					3	5			2	2
Tool and Object Nomenclature, Use and Recognition	2		1	1	2	2								1			5	6
Dual Task: construction score	3	1		2	3	4	1				1	1			2			5
Dual Task: watch score			6	3	5									4				
Dual Task: latency score		5								4					4			4
Coordinative Speed and Accuracy: speed and accuracy score			2				2	1	1	1		2			3			
Coordinative Speed and Accuracy: accuracy score							4			6		2	3	4		6		
Level of Aspiration: initial goal estimate		3								2	6	4		3		6		
Level of Aspiration: realism						5		5	5									
Group Task		6	5	6		3						5	1	6	1	3	1	1
Interview: mean score							6	3	2			6	2			1	3	
Interview: items not rated	4										5					5		
Multiple R based on optimum combination of test scores	.84	1.00	1.00		.95	.91 .85	.77	1.00	1.00		.92	.91 .82	.66	1.00	1.00		.85	.77 .77
N	140	13	15		26	44 42	140	13	15		26	44 42	140	13	15		26	44 42

The exercise most emphasized by the policy of each administrator and the three exercises most emphasized by the group policy were all designed to assess the learning dimension. However, some exercises designed to tap other dimensions were used in the learning dimension policies of both the individuals and the group. For example, for administrator number 1, the score entered third into the stepwise multiple regression analysis (and, accordingly, emphasized more than all the other scores but two) was a Level of Aspiration score. The Level of Aspiration score was designed to assess the Social/Motivational dimension. Other social/motivational scores were included by other administrators. The inclusion of social oriented test scores in the policy for the learning dimension might be an indication of a "halo effect." The halo effect in this case would be an unwarranted carryover of a high rating in one dimension to the rating of another dimension. However, analysis of the social related test scores included in the policy relative to the learning dimension indicated that the weighting of these scores was small and sometimes negative. Further, the overall correlation of the ratings on the learning and social dimensions was .27, which indicates only a modest relationship between the two ratings.

Policy in Rating Men on Psychophysical-Motor Dimension

The individual multiple R values for administrator policy capturing regression analyses relative to the psychophysical-motor dimension were: 1.00, 1.00, .92, .91, and .82. Following the logic presented earlier, this suggests acceptable intrarater agreement.

The combined policy of all administrators in the psychophysical-motor dimension was investigated in the same manner as for the learning dimension. The results of this analysis indicated that the group policy emphasized the Dual Task construction score and a Coordinative Speed and Accuracy score (Table 16). All of the four most heavily weighted scores were designed to evaluate this dimension. The obtained multiple R was .77.

The policies for the individual administrators were, as expected, more fully captured than the group policy. The individual policies are summarized in Table 15. The score most emphasized by each administrator was one of the two most emphasized by the group policy (described above) and was a test designed to assess the psychophysical and motor dimension.

Table 16

Group Policy for Rating Men on the Psychophysical-Motor Dimension

Step	Score	Stepwise Multiple R	R^2	Increase in R^2	Zero Order r
1	Dual Task: construction score	.61	.38	.38	.61
2	Coordinative Speed and Accuracy: speed and accuracy score	.69	.48	.10	.56
3	Inspection-Sort: speed and accuracy score	.73	.53	.05	.40
4	Coordinative Speed and Accuracy: accuracy score	.74	.55	.02	.39
5	Reliability	.75	.56	.01	-.20
6	Interview: mean rating	.75	.56	.01	.03

Multiple R based on optimum combination of scores= .77

However, the policies of both the group and the individuals included some emphasis on scores intended to assess other dimensions. The extent of this cross-dimensional influence can be estimated by the correlation between the psychophysical and motor dimension and the other dimensions. These correlations, as stated above, are .42 and .25 for the learning and the social/motivational dimensions, respectively. These are modest levels and indicate a relatively large independence of these dimensions.

Policy in Rating Men on Social/Motivational Dimension

The combined policy of the TCAC administrators on the social/motivational dimension was less uniform than the combined policies on the other two dimensions. There was moderately high consistency in the policy of each administrator on the social/motivational dimension. The multiple Rs indexing this are: 1.00, 1.00, .85, .77, and .77. The analysis is summarized in Table 15.

The two scores most emphasized by the policies of the individual administrators were not designed to assess the social/motivational dimension. This indicates that the administrators had some difficulty in making these ratings using the exercise scores. However, the overlap of the ratings of this dimension with those of the other two dimensions was low, as indicated by the correlation coefficients of .27 and .25 with the learning and the psychophysical

and motor dimensions respectively. The overall agreement of the individual policies is reflected in the multiple R between the test scores and the ratings of all administrators. This multiple R was .66. The three tests most emphasized by the group policy were all social/motivational related tests: the Group task, the Interview mean score, and the Reliability exercise. This analysis is summarized in Table 17.

Table 17

Group Policy for Rating Men on the Social/Motivational Dimension

Step	*Score*	*Stepwise Multiple R*	R^2	*Increase in* R^2	*Zero Order r*
1	Group Task	.51	.26	.26	.51
2	Interview: mean score	.56	.31	.05	.28
3	Reliability	.60	.36	.05	-.26
4	Coordinative Speed and Accuracy: accuracy score	.62	.38	.02	.27
5	Inspection/Sort: accuracy score	.63	.39	.01	-.12
6	Conceptual Integration/Application	.63	.40	.01	.20

Multiple R based on optimum combination of Center test scores= .66

TCAC Career Field Recommendations

The purpose of the TCAC procedures was to reclassify low aptitude enlistees and, where possible, to identify career fields in which they might do well. The resulting classifications were summarized earlier (Table 6).

The career field recommendations of the Center were based on a vote which followed a group discussion in which all data for an individual were evaluated. The extent to which the career field recommendations of the TCAC can be predicted directly from the various scores is of interest. This form of policy capturing has

been called "paramorphic representation" of the judgmental process by a mathematical model (Dawes, 1971). This approach possesses important implications for large scale application of the TCAC method because it investigates the potential of replacing the group decision making process by its mathematical representation. This potential approach has also been termed "bootstrapping." Quite suprisingly, when criterion data are available, this technique has, in a variety of situations, been shown to be more effective (valid) than the performance of the judges on which the bootstrapping model is based. Alternative explanations for this emphasize either: (a) the ability of the model to detect the true policy and reject unwanted noise, and/or (b) the relative robustness of linear models (Dawes & Corrigan, 1974).

The extent to which the career field recommendations can be predicted by the various scores was investigated using a discriminant analysis technique. This technique can be used to classify individuals based on their relative similarity to each of two or more subgroups: in the current case, the actual classification recommendation for general rates or for a specific rating. The analysis indicated that the means of the various test scores of individuals receiving the two types of recommendations were significantly different (generalized Mahalanobis' $D^2 = 71.3$, which is distributed as chi square with 16 d.f., $p < .05$). The results of the classification analysis are summarized in Table 18. The classifications based on this analysis are 78.6 percent accurate. A linear function of the test scores thus captures well the policy used by the TCAC administrators in making career field recommendations. Based on the discriminant analysis, it appears that the bootstrapping approach holds promise for considerably streamlining the TCAC procedures.

Table 18

Comparison of Actual and Statistical Classifications of Men

Discriminant Analysis	Actual Recommendation: General Rates	Actual Recommendation: Specific Rating	Total No. of Men
General Rates	47	13	64
Specific Rating	17	63	76
			140

Generalized Mahalanobis' $D^2 = 71.3$, $p < .05$ (distributed as chi square with 16 d.f.)

Percentage correct classifications= 78.6.

Prediction of On-the-Job Success

The decisions of the TCAC concerning the level of success expected in the Fleet were clinical (subjective) in nature. These decisions were, however, based on and cast against the backdrop of the data which evolved from the various procedures. The approach applied in this analysis attempts to make explicit, through multiple regression analysis, those tests emphasized by the TCAC administrators in reaching decisions concerning predicted level of success in the field. Additionally, the utility of the Navy test scores in explaining (i.e., capturing the policy of) the TCAC predictions of job success was investigated.

Initially, predictions of job success were thought to be dependent on the preceding career field recommendation made by the Center. As described above, the discriminant analysis indicated that the two subgroups differed significantly with respect to the mean scores on the TCAC exercises. Therefore, analyses were completed separately for men recommended for general rates and men recommended for machinist's mate. Only these two subsets had a sufficiently large number to justify analysis. Additionally, an analysis of the total sample was completed.

Group Policy for Machinist's Mate and General Rates Subsamples

Two multiple linear regression analyses identified the linear combination of TCAC test scores that best predicted the Center predictions of on-the-job performance for two subsamples of men: those recommended for machinist's mate (N = 38), and those recommended for general rates (N = 64). The six scores that together accounted for the greatest amount of variance in the predicted level of success for each of these two subgroups are given in Table 19 for the machinist's mate subsample and in Table 20 for the general rates subsample. For the machinist's mate subsample, the multiple correlation, based on 16 center measures, was .90. For the general rates subsample, the multiple correlation, based on 16 center measures, was .70. The basic interpretation of these two analyses is that the policy of the Center in predicting Fleet job success was substantially "captured" for both subsamples.

Table 19

Policy Capturing of Level of Job Success in the Fleet by Score--Machinist's Mate Subsample (N = 38)

Step	*Score*	*Stepwise Multiple R*	R^2	*Increase in* R^2	*Zero Order r*
1	Reliability	.53	.28	.28	-.53
2	Conceptual Integration/Application	.65	.42	.14	.50
3	Coordinative Speed and Accuracy: speed score	.72	.52	.10	.29
4	Group Task	.77	.59	.07	.36
5	Tool and Object Nomenclature, Use, and Recognition	.79	.63	.04	.09
6	Coordinative Speed and Accuracy: accuracy score	.81	.65	.02	.37

Multiple R based on optimum combination of scores= .90

Table 20

Policy Capturing of Level of Job Success in the Fleet by Score--General Rates Subsample (N = 64)

Step	Score	Stepwise Multiple R	R^2	Increase in R^2	Zero Order r
1	Conceptual Integration/Application	.47	.22	.22	.47
2	Reliability	.57	.32	.10	-.25
3	Interview: mean score	.64	.40	.08	.31
4	Dual Task: construction score	.66	.43	.03	-.10
5	Inspection/Sort: accuracy - good items	.67	.45	.01	-.22
6	Inspection/Sort: speed and accuracy	.68	.46	.01	.05

Multiple R based on optimum combination of scores= .70

The Reliability and the Conceptual Integration/Application scores were identified by these analyses as the most important factors in decisions concerning both subsamples. The emphasis of the administrators on the Reliability test can be further explained on the basis of observations made at meetings at which these predictions and the recommendation of ratings were made. It appears that the Reliability exercise score was used as a "killer" item. The staff generally thought poorly of those men who did poorly on the Reliability exercise. These men were, generally, not recommended to specific ratings. However, two men recommended to be machinist's mates failed this test. They possessed, presumably, other positive attributes that caused the staff to recommend them. These men were predicted to have only minimally acceptable level of on-the-job success.

The second exercise identified as a primary constituent of the prediction policy, the Conceptual Integration/Application, involves a complex troubleshooting task. This exercise was designed to assess the learning dimension. This score was most highly correlated with level of success predictions for the general rates subsample (zero order correlation = .47). It was the second most highly correlated score with the level of success prediction for the machinist's mate subsample (zero order correlation = .50). We note that the two subsamples differed significantly in mean level of performance on this test ($t = 3.0$, $p < .05$). This exercise score, then, was an important part of the policy of the TCAC staff's level of on-the-job success predictions within both of these subgroups.

For both subgroups, these two exercises (Reliability and Conceptual Integration/Application) largely describe the policy of the administrators with respect to the predicted level of job success within the two subgroups involved. The other scores added incrementally and significantly. However, no single test alone stands out as particularly or predominantly important in the formation of the rating policy. There was no overlap between the subgroups in the next four tests identified by the multiple regression analysis as adding most to the predictable variance of level of success. For each subgroup, the first six tests that are identified by the multiple regression analysis include tests designed to assess each of the three dimensions included.

Group Policy With Respect to the Total Sample

The predictions of level of job success may be expected to be highly dependent on the prior Center recommendation of a specific career field. However, it is heuristic to assess the Center's policy in making these predictions for the total sample, disregarding temporarily the previously mentioned problems introduced by career field considerations.

A stepwise multiple linear regression analysis identified the linear combination of scores that best captured the staff's policy for the total group. The six scores that together accounted for the highest amount of variance in predicted scores are given in Table 21. Tests designed to assess all three dimensions are represented in these six tests. The complex Conceptual Integration/Application score was most heavily weighted. This emphasis

reflects the staff's emphasis on the learning dimension. The Reliability score was next most highly influential in predicting the on-the-job success. This exercise reflects the social/motivational dimension.

Table 21

Policy Capturing of Level of Job Success in the Fleet by Score--Total Sample (N = 140)

Step	*Score*	*Stepwise Multiple R*	R^2	*Increase in* R^2	*Zero Order r*
1	Conceptual Integration/ Application	.47	.22	.22	.47
2	Reliability	.58	.23	.11	-.33
3	Interview: mean rating	.61	.37	.04	.27
4	Group Task: total score	.63	.40	.02	.24
5	Coordinative Speed and Accuracy: speed score	.64	.41	.01	.27
6	Inspection/Sort: accuracy score	.65	.42	.01	-.05

Multiple R based on optimum combination of scores= .68

Overall, the six most important test scores identified by the multiple regression analysis accounted for 42 per cent of the variance of the administrator's predictions of fleet job success (multiple R = .68). If one interprets the group decision as a criterion, this level of prediction is quite impressive compared to other predictive validity studies.

The policy concerning the total sample can be compared to the policies for each of the two subgroups previously discussed: men recommended for machinist's mate and men recommended for general rates. This comparison can be made first concerning the extent of policy capturing, and second concerning the tests identified by the multiple regression analysis as important. As would be anticipated, the policy for the total group overlaps considerably with the policies of the two subgroup constituent parts of the total group. The two exercises most emphasized in the three policies were identical; the Conceptual Integration/Application and the Reliability tests. The remaining four of the six most influential scores in the whole group policy include two tests emphasized by each of the two subgroup policies. In all, the six tests most emphasized by the total group policy included tests from each of the three dimensions identified by the job analysis.

The extent to which the variance of the policy was captured, as indexed by the multiple R based on all 16 Center measures, is .68 for the whole group, .90 for the machinist's mate subgroup, and .70 for the general rates subgroup. The shrinkage in the multiple R from the machinist's mate subgroup to the whole sample is understandable in terms of: (a) the expected more stringent requirements set by the Center for a given level of success prediction for men recommended for more demanding ratings, and (b) possible variations in policy for the eight types of career field recommendation subgroups depending upon the dimensional requirement of the specific ratings.

Comparison of Navy and TCAC Scores

Analyses were conducted to determine the extent to which the Navy test scores (GCT, ARI, MECH, CLER, ETST, and SHOP PR) could be used to explain (capture the policy of) the level of the job success predictions. These scores were subjected to the stepwise multiple regression analyses as reported above for the Center exercise scores. The multiple correlation coefficients obtained were .12, .41, and .23 for the general rates subgroup, machinist's mate subgroup, and the whole sample respectively. The corresponding multiple Rs reported above based on six Center test scores are, respectively, .68, .81, and .65. Accordingly, the Navy tests can be seen to have relatively little utility in capturing the policy of the TCAC relative to level of success predictions. Further, their utility varies relatively greatly from subgroup to subgroup, and from

subgroups to the whole sample. This is quite understandable as the TCAC exercises were designed to be the basis for the staff's decisions. These comparisons further indicate that the decisions of the TCAC administrators were based on the Center exercise scores and that these scores are not redundant with the available Navy test scores. These analyses are summarized in Tables 22, 23, and 24.

Table 22

Policy Capturing of Level of Job Success in the Fleet by Navy Test Score--General Rates Subsample (N = 64)

Step	*Navy Test*	*Stepwise Multiple R*	R^2	*Increase in* R^2	*Zero Order r*
1	SHOP PR	.09	.01	.01	.09
2	MECH	.11	.01	0*	.03
3	ETST	.12	.01	0*	.02

Multiple R based on optimal combination of Navy test scores= .12
**less than .005*

Table 23

Policy Capturing of Level of Job Success in the Fleet by Navy Test Score--Machinist's Mate Subsample (N = 38)

Step	*Navy Test*	*Stepwise Multiple R*	R^2	*Increase in* R^2	*Zero Order r*
1	CLER	.28	.08	.08	-.28
2	SHOP PR	.35	.12	.04	.28
3	ETST	.39	.15	.03	-.05
4	GCT	.40	.16	.01	.13
5	ARI	.41	.17	0*	-.16

Multiple R based on optimal combination of Navy test scores= .41
**less than .005*

Table 24

Policy Capturing of Level of Job Success in the Fleet by Navy Test Scores--Total Sample (N = 140)

Step	*Navy Test*	*Stepwise Multiple R*	R^2	*Increase in* R^2	*Zero Order r*
1	SHOP PR	.16	.03	.03	.16
2	ETST	.19	.03	.01	-.05
3	MECH	.21	.04	.01	.14
4	CLER	.22	.05	0*	.04
5	ARI	.22	.05	0*	.02
6	GCT	.23	.05	0*	.02

Multiple R based on optimal combination of Navy test scores= .23
**less than .005*

Views of Individuals Assessed

The opinions of the individuals assessed in the TCAC relative to the various exercises were systematically elicited through a questionnaire which contained both multiple choice and open-ended questions. The questionnaire was administered at the conclusion of the assessment day.

The multiple choice data concerning the perceived fairness indicated unanimous agreement that the TCAC exercises were fair in an absolute sense. In a choice between "yes" and "no," indicating agreement or disagreement with a statement that the TCAC tests seem fair, 100 per cent of the sample responded "yes." This topic was further probed by asking the individuals who participated in the assessment process to compare the TCAC exercises with other tests taken in the past. Based on a five category Likert type scale, 31 per cent of the enlistees considered the TCAC exercise tests "very much more fair," 42 per cent found the TCAC exercises "more fair," and 26 per cent said the TCAC exercises were "equally fair." Only one individual (less than 1 per cent of the sample) chose

the "less fair" category. No one chose the "very much less fair" category. There was also unanimous agreement that the TCAC exercises were enjoyable. On a four category scale, 76 per cent of the individuals assessed chose the superlative "enjoyed very much" response. No one chose the nonenjoyable category.

Reasons for Fairness Evaluation

The reasons for perceiving the exercises to be fair were inquired into through a completion type question. Typical replies to this question are listed below. Reasons frequently given for thinking the TCAC exercises were fair fell in two major areas: (a) the training completeness and hands on aspects of the learning phase of the exercises, and (b) the exercises emphasized performance, not reading and writing.

"Gave me a chance to prove that I could do some things with my hands, not just my head."

"Very fair. It was explained very well every time, to make sure I understood what was going on, and I did."

"Yes. Because these weren't tricky like others."

"It gives a person an opportunity to use his hands in a certain situation, instead of learning it from a book. Like some tests you have to read questions and a certain amount of time is alloted (sic) for the test. Some people can't read fast enough and that isn't fair. But here you can use your hands."

"Because it was a lot less writing and more doing."

Opinions Concerning Specific Exercises

The questionnaire asked the individuals who were assessed to indicate (using a checkoff list) the two exercises they "liked *most*" and the two they "liked least." A number of times the persons processed volunteered that there were no tests that they "disliked" but that they chose the two tests that they liked least to respond to the question. The three exercises liked best were the: Dual Task, Conceptual Integration/Application, and the Coordinative Speed and Accuracy. These were all job learning performance exercises. This agrees with the stated opinions that the exercises were appreciated which taught requisite skills and allowed performance of actual tasks. Additionally, there was some emphasis in response to this question relative to the "challenge" offered by these exercises. The number of individuals choosing a test as "liked best" correlated negatively (rho = .40) with the number of enlistees who chose that test as "liked least." The three exercises liked least were: Reliability, Level of Aspiration, and Group Task. There was no one unifying reason for these being "liked least." However, relative to the Group Task, several persons mentioned that they did not like performing in a group before they knew the other group members.

DISCUSSION AND CONCLUSIONS

The major purposes of the present program, established at the outset, were to develop and demonstrate a fair classification/reclassification procedure for enlisted personnel in the Navy and to demonstrate the developed procedures. To these ends, two recent concepts were applied to provide the basic foundation for the procedures. The first concept was the assessment center approach. This concept, which has become increasingly accepted in managerial potential evaluation, has been only seldomly employed in the technical performance potential evaluative sphere. The second concept was based on a test development model which contends that the person who demonstrates the ability to learn a sample of a job in a training and evaluative situation will, given appropriate job training, be able to learn the total job.

This latter concept represents a novel test development point of view and represents a departure from orthodox test development models. While such a concept is not believed to be a substitute for predictive evaluation, the concept is believed to represent a practical approach to assuring commonality between predictor and criterion. Certainly, the concept seems to possess more common sense than the approach, often employed, which attempts to demonstrate that predictor scores correlate well with some other concurrent measures. Two concurrent measures can be highly correlated, but both may possess no long term predictive validity. In this case, the concurrent validity concept seems meaningless. We suggest that the concurrent criterion approach often receives greater emphasis and is of less value than most test constructors seem willing to admit. Consider school grades, an often used (misused?) criterion employed to establish the "validity" of a test. The correlation between the predictor and school grades is established by the test constructor. The correlation between school grades and job success is known to be low both in general and in the Navy situation. With imperfect predictor-school grade correlation and with imperfect school grade-job performance correlation, there seems to be a degree of fuzzy thinking relative to defending such a paradigm relative to predictor-job performance validity (where predictive validity is defined in terms of predicting ultimate job performance). The present demonstration of the ability to learn the job seems more defensible and straightforward.

Preceding page blank

Relative to the predictive validity issue, we do not contend that the assessments produced by the TCAC need not be validated. While the results of the present study indicated considerable reliability and internal consistency, they tell us little about the actual performance in the selected career fields of the men evaluated. A Fleet followup study is being conducted in this regard and seems especially important in view of the other positive aspects found for the TCAC approach.

TCAC was indicated to be viable with respect to all areas investigated. Overall, the procedures were practical to implement and were well received by the individuals assessed. They perceived the tests as both fair and agreeable. In particular, the job learning aspect of many of the tests was well received. The performance measures were adequately reliable and the policies followed by the administrators were consistently applied. The policies of the several administrators were in close agreement with one another. The exercise scores have only modest overlap with currently used Navy tests. As a prototype, the TCAC, as implemented, seemed quite acceptable. However, areas for improvement exist. For example, a level of aspiration exercise which is more Navy job oriented seems desirable. The acceptance of the Reliability exercise also seems to have suffered because of low perceived job applicability.

It has been contended by others that the performance approach is uneconomical because of its imposition of excessive participant time requirements. To this we answer that six hours were required to administer all TCAC exercises to each group. Administration of the usual Navy tests requires about four hours. In view of the potential power of the TCAC approach, the additional two hours required by the performance approach does not seem excessive. Moreover, the performance approach need not be administered to all enlisted personnel. Those who achieve an A school recommendation by the usual classification procedures do not require extra concern. However, those who might in fact do well in a career field and do not receive such acknowledgment from the usual procedures certainly deserve a "second chance." The TCAC approach could provide this second chance.

The preferred composition of the evaluative team remains open. The present program's administrator team was composed of two psychologists and two Navy Chief Petty Officers. It was anticipated that the psychologists would provide appropriate personality-social/motivational appraisal and test administrative expertise.

On the other hand, the Chief Petty Officers were anticipated to provide job detail/career field expertise not possessed by psychologists. Yet, the data indicated all four administrators to be employing consistent policies, i. e., similarly relying on and weighting the exercise scores to derive recommendations. The assessor team may have become quickly homogenized or the various points of view may not have been fully represented by the assessors. Alternatively, the anticipated points of view may have been involved subtly but not reflected by the quantitative procedures.

There was also some indication that at least a part of the group decision making could be replaced by a mathematical technique. This would further streamline the process.

The TCAC was designed with culture fairness in mind. To this end, oral and performance exercises were emphasized and hands on training was implemented. There was minimum emphasis on reading skills.

Finally, we do not contend that the single "best" individual or combination of exercises was developed. While substantial merit was demonstrated for the exercises employed, alternative or improvements in the present exercises are entirely possible. Prototypes by their very nature serve to prove a concept. Full production models often depart from design specifics incorporated within a prototype.

The data presented throughout the body of this report warrant the following conclusions:

- The TCAC concept, including the job learning-testing and multiple assessment aspects, represents a potentially useful and workable classification/reclassification approach.
- The exercises and methods included in the TCAC were such as to allow acceptable within and between administrator reliability.
- The individual administrators each showed high consistency in making judgmental evaluative decisions.

- Each of the three job analytically based dimensions--learning, psychophysical-motor, and social/motivational--considered in deriving career field recommendations was relatively independent and useful for achieving the required classifications.

- A consistent policy was employed by the TCAC in forming career field recommendations and in predicting job performance success.

- Assuming adequate evidence of predictive validity, the TCAC concept could make a significant contribution towards increased effectiveness of personnel utilization in the Navy.

REFERENCES

Albright, L.E., Glennon, J.R., & Smith, W.J. *The use of psychological tests in industry.* Cleveland: Howard Allen, Inc., 1963.

Adkins, D.C., Primoff, E.S., McAdoo, H.L., Bridges, C.F., & Forer, B. *Construction and analysis of achievement tests: The development of written and performance tests of achievement for predicting job performance of public personnel.* Washington, D.C.: U.S. Goverment Printing Office, 1947.

American Psychological Association, *Standards for educational and psychological tests.* Washington, D.C.: American Psychological Association, 1974.

Ash, P., & Krocker, J.F. Personnel selection, classification, and placement. In M.R. Rosenberg and L.W. Porter (Eds.), *Annual Review of Psychology,* 1975, *26,* 481-508.

Bornstein, H., Jensen, B.T., & Dunn, T.F. The reliability of scoring in performance testing as a function of tangibility of the performance product. *American Psychologist,* 1954, *9,* 336-337.

Bottenberg, R.A., & Christal, R.E. Grouping criteria: A method which retains maximum predictive efficiency. *Journal of Experimental Education,* 1968, *4,* 28-34.

Bray, D.W., & Campbell, R.J. Selection of salesmen by means of an assessment center. *Journal of Applied Psychology,* 1968, *52,* 36-41.

Bray, D.W., & Grant, D.L. The assessment center in the measurement of potential for business management. *Psychological Monographs,* 1966, *80,* no. 17, whole no. 625.

Bray, D.W., & Moses, J.L. Personnel selection. *Annual Review of Psychology,* 1972, *23,* 545-576.

Campion, J. Work sampling for personnel selection. *Journal of Applied Psychology,* 1972, *56,* 40-44.

Christal, R.E. *JAN: A technique for analyzing group judgment.* Lackland Air Force Base, Texas: Personnel Research Laboratory and Aerospace Medical Division, February 1963. (PRL-TRD-63-3, ASTIA Document AD-403813).

Christal, R.E. JAN: A technique for analyzing group judgment. *Journal of Experimental Education,* 1968, *4,* 24-27. (a)

Christal, R.E. Selecting a harem - and other applications of the policy-capturing model. *Journal of Experimental Education*, 1968, *4*, 35-41. (b)

Dawes, R.M. *Graduate admissions: A case study*. Eugene, Oregon: Oregon Research Institute, 1970.

Dawes, R.M. A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist*, 1971, *26*, 180-188.

Dawes, R.M., & Corrigan, B. Linear models in decision making. *Psychological Bulletin*, 1974, *81*, 95-106.

First court test of assessment center method. *Assessment and Development*, 1976, *3*, 1-2.

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

Ghiselli, E.E. The validity of aptitude tests in personnel selection. *Personnel Psychology*, 1973, *26*, 461-477.

Goldberg, L.R. Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, 1970, *73*, 422-432.

Greenwood, J.M., & McNamara, W.J. Interrater reliability in situational tests. *Journal of Applied Psychology*, 1967, *51*, 101-106.

Greenwood, J.M., & McNamara, W.J. Leadership styles of structure and consideration and managerial effectiveness. *Personnel Psychology*, 1969, *22*, 141-152.

Greer, F.L., Pearson, W.G., & Havron, W.O. *Evasion and survival problems and the prediction of crew performance*. Lackland Air Force Base, Texas: Office for Social Science Programs, Air Force Personnel and Training Research Center, Air Research and Development Command, 1957.

Guilford, J.P. *Fundamentals of statistics in psychology and education* (4th ed). New York: McGraw-Hill, 1965.

Guilford, J.P., & Hoepfner, R. *The analysis of intelligence*. New York: McGraw-Hill, 1971.

Hammond, K.R., Hursch, G.J., & Todd, F.J. Analyzing the components of clinical inference. *Psychological Review*, 1964, *71*, 438-456.

Hinrichs, J.R. Comparison of "real life" assessments of management potential with situational exercises, paper-and-pencil ability tests, and personality inventories. *Journal of Applied Psychology*, 1969, *53*, 425-432.

Holden, C. Employment testing: Debate simmers in and out of court. *Science*, 1975, *190*, 35-39.

Howard, E.A. An assessment of assessment centers. *Academy of Management Journal*, 1974, *17*, 115-134.

Huck, J.R. Assessment centers: A review of the external and internal validities. *Personnel Psychology*, 1973, *36*, 191-212.

Jensen, J.J., Hill, J.H., Siegel, A.I., & Courtney, D. *The development of a job sample troubleshooting examination.* Philadelphia: Institute for Research in Human Relations, 1954.

Lifter, M. Validation requirement clarified by supreme court. *The Industrial-Organizational Psychologist*, 1976, *13*, 24-26.

MacKinnon, D.W. *An overview of assessment centers. Technical report no. 1.* Berkeley, Ca.: Center for creative leadership, University of California, 1975.

Madden, J.M. *A policy capturing model for analyzing individual and group judgment for job evaluation.* Lackland AFB, Texas: 6570th Personnel Research Laboratory, Aerospace Division, 1964.

McCormick, E.J., & Tiffin, J. *Industrial psychology.* (6th. edition). Englewood Cliffs, New Jersey: Prentice Hall, 1975.

Mitchell, J.O. Assessment center validity: A logitudinal study. *Journal of Applied Psychology*, 1975, *60*, 573-579.

Moses, J.L., et al, *Standards and ethical considerations for assessment center operations.* Quebec: Task Force on Development of Assessment Center Standards, Third International Congress on the Assessment Center Method, 1975.

NAVPERS 18068C: Manual of qualifications for advancement in rating. Washington, D.C.: Department of the Navy, Bureau of Personnel, 1971.

O'Leary, L.R. Employment, sound psychometric practice, and reality. *American Psychologist*, 1973, *28*, 147-150.

Rundquist, E.A. The prediction ceiling. *Personnel Psychology*, 1969, *22*, 109-116.

Siegel, A.I., & Bergman, B.A. *Nonverbal and culture fair performance prediction procedures: I. Background, test development and initial results.* Wayne, Pa.: Applied Psychological Services, 1972.

Siegel, A.I., & Bergman, B.A. A job learning approach to performance prediction. *Personnel Psychology*, 1975, *28*, 325-339.

Siegel, A.I., Bergman, B.A., & Lambert, J.V. *Nonverbal and culture fair performance prediction procedures: II. Initial validation.* Wayne, Pa.: Applied Psychological Services, 1973.

Siegel, A.I., & Courtney, D. *Development of practical performance measures.* Philadelphia: Institute for Research in Human Relations, 1953.

Siegel, A.I., & Leahy, W.R. *Nonverbal culture fair performance prediction procedures. III. Cross validation.* Wayne, Pa.: Applied Psychological Services, 1974.

Siegel, A.I., & Federman, P.J. *Policy capturing for deriving content emphasis in technical training.* Wayne, Pa.: Applied Psychological Services, in press.

Slovic, P. Analyzing the expert judge: A descriptive study of a stock-broker's decision processes. *Journal of Applied Psychology*, 1969, *53*, 255-263.

Slovic, P., & Lichtenstein, S. *Comparisons of Bayesian and regression approaches to study of information processing in judgment.* Eugene, Oregon: Oregon Research Institute, 1970.

Stephenson, R.W., & Ward, J.H. Computer-assisted discussions to help a policy group assign weights to criterion ratings. *Personnel Psychology*, 1971, *24*, 447-461.

Supreme Court of the United States. *Case of Willie S. Griggs, et al. VS Duke Power Company.* Washington, D.C.: Bureau of National Affairs, 1971.

Tiffin, J. *Industrial psychology (3rd. edition).* New York, Prentice-hall, 1952.

Wiesen J.P., & Siegel, A.I. *Administrator's manual for the Navy Experimental technical assessment center.* Wayne: Pa.: Applied Psychological Services, 1975.

Wollowick, H.B., & McNamara, W.J. Relationship of the components of an assessment center to management success. *Journal of Applied Psychology*, 1969, *53*, 348-352.