227927

UNCLASSIFIED

RESEARCH ON THE DEVELOPMENT OF SHIPBOARD PERFORMANCE MEASURES

Technical Report IX

SUPERVISORY RATINGS AND PRACTICAL PERFORMANCE TESTS AS COMPLEMENTARY CRITERIA OF SHIPBOARD PERFORMANCE

Robert R. Mackie, Ph.D.
and
Wallace S. High, M.A.

Prepared for

Psychological Sciences Division
Personnel and Training Branch
Office of Naval Research
Department of the Navy

by

Human Factors Research, Incorporated
Los Angeles, California

June 1959

Contract No. Nonr 1241(00)

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

UNCLASSIFIED

ABSTRACT

This report describes a shipboard follow-up study of the performance of Navy Machinery Repairmen whose aptitudes, skills, interests and achievements had been thoroughly studied two years earlier while they were in Class "A" MR training.

Shipboard performance was assessed in two ways: (1) by administering a practical performance test requiring skill in the use of machinery repair equipment; (2) by securing ratings by supervising petty officers of each person's ability to perform the various aspects of the MR's shipboard job.

The results strongly suggested that performance tests and supervisory ratings were best regarded as complementary criteria of shipboard performance. While these two measures did not correlate with each other, both correlated significantly with many logical predictors, including aptitude and interest measures, practical work during training, and predictions of success by Class "A" School instructors.

When the two shipboard measures were combined to form a simple composite criterion, it was estimated that over 50% of the true variance was accounted for by scores made two years earlier on a combination of scores made on: (1) mechanical knowledge tests; (2) training projects involving the use of lathe and milling machines; and (3) predictions by school instructors as to eventual suitability as an MR.

UNCLASSIFIED

ACKNOWLEDGEMENTS

The cooperation of many individual commands, and their engineering officers and petty officers, was instrumental to the sucess of this study:

USS Abbot (DD-629)
USS Ajax (AR-6)
USS Ammen (DD-527)
USS Amphion (AR-13)
USS Belle Grove (LSD-2)
USS Bolster (ARS-38)
USS Bronson (DD-668)
USS Cadmus (AR-14)
USS Compton (DD-705)
USS Comstock (LSD-19)
USS Dixie (AD-14)
USS Fulton (AS-11)
USS Grand Canyon (AS-28)
USS Gyatt (DDG-1)
USS H. R. Dickson (DD-708)
USS Hector (AR-7)
USS Hyman (DD-732)
USS Jason (AR-8)
USS John S. McCain (DL-3)
USS Kermit Roosevelt (ARG-16)
USS Moale (DD-693)
U S Naval Air Station (Point Mugu)
USS Norris (DDE-859)
USS Orion (AS-18)
USS Piedmont (AD-17)
USS Putnam (DD-757)
USS Rogers (DDR-876)
USS Rankin (AKH-103)
USS Rushmore (LSD-14)
USS Salisbury Sound (AV-13)
USS Shadwell (LSD-15)
USS Sperry (AS-12)
USS Taconic (AGC-17)
USS Yellowstone (AD-27)
USS Yorktown (CVS-10)

Special appreciation is extended to Chief John Justyna (MRC) for his help in designing the test product and to the Commanding Officer, Service School Command, NTC, San Diego, for his continued cooperation in making research studies possible within the Class "A" Schools.

TABLE OF CONTENTS

UNCLASSIFIED

CHAPTER		Page

UNCLASSIFIED

CHAPTER Page

UNCLASSIFIED

LIST OF TABLES AND FIGURES

FIGURE		Page
1	Shipboard performance test blueprint	11
2	Sample page from shipboard performance rating scale	16

SUPERVISORY RATINGS AND PRACTICAL PERFORMANCE TESTS AS COMPLEMENTARY CRITERIA OF SHIPBOARD PERFORMANCE

Chapter 1

SUMMARY AND OPERATIONAL IMPLICATIONS

BACKGROUND

This was one of a series of studies designed to investigate the utility of various measures that might be employed as criteria of shipboard performance. Previous studies in this series had indicated that both job sample tests and supervisory rating scales could be developed that would yield reliable and discriminating measures of the shipboard performance of Navy enlisted personnel (10,11,12,13,14,15). However, these studies indicated very little common variance between performance tests and supervisory ratings. This lack of correlation created questions concerning the extent of valid variance in both types of criterion measures. The present study was designed to examine more completely the relationships between supervisory ratings and performance tests given aboard ship and to describe their nature by examining their variance in common with a substantial number of other measures.

PROCEDURE

A job sample performance test for Machinery Repairmen requiring the construction of a specially designed product through the use of shipboard lathes, shapers, drill presses, and hand tools was developed together with a specially tailored rating scale covering both the technical and non-technical aspects of the MR's shipboard job. The subjects for study were 64 Machinery Repairmen serving aboard Navy destroyers, aircraft carriers, tenders, and repair ships. These men had been the subject of study two years earlier while undergoing Class "A" MR training during which time extensive measures of their basic aptitudes, interests,

UNCLASSIFIED

and school achievement were made. This study investigated the relationships between the measures obtained in the school environment and those obtained two years later in the operating environment.

RESULTS

The relative independence of scores on performance tests designed to sample the practical aspects of the shipboard job and scores on rating scale traits also defined to reflect the various aspects of the job was again observed. The correlation between the total score on the shipboard job performance test and average score on the supervisory rating scale for the 64 subjects was .03.

In spite of this independence, both shipboard proficiency measures showed substantial relationships to various measures obtained two years earlier in the training environment. Ten of the 20 predictor measures used correlated significantly with one or both of the shipboard measures.

Scores on the job sample performance test administered aboard ship correlated:

- .44 with scores made during Class "A" training on school work samples requiring basic milling, fine lathe threading, and hand tool performance skills;
- .39 with composite scores made during Class "A" training on tests of mechanical knowledge, perceptual speed, and spatial relations;
- .32 with predictions (ratings) made by school instructors two years earlier on suitability for doing the job of a Machinery Repairman.

Scores from the supervisor's rating scale administered aboard ship also correlated significantly with logical predictors from the school environment:

- .44 with scores derived from school work samples in the Class "A" training environment requiring skills in lathe turning to specifications and fine lathe treading;

.29 with scores on a test of mechanical knowledge administered at the time of Class "A" training;

.35 on measures of achievement in MR School as reflected by scores on written examinations for each phase of training;

.35 with predictions (ratings) made by the MR School instructors on the suitability of each graduate for doing the job of a Machinery Repairman.

Since both the shipboard job performance test scores and the shipboard supervisory rating scores correlated significantly with a number of logical predictors and yet were essentially independent of each other, a simple unweighted composite of the two measures was formed. This composite shipboard performance criterion correlated:

.51 with scores made on school work samples requiring skill in fine lathe threading and basic milling machine operations;

.47 with intellectual aptitude factors commonly defined as mechanical knowledge, spatial relations, and perceptual speed;

.27 with interest in mechanical rates as measured by the Clark Navy Interest Profile;

.35 with school achievement in Class "A" training as measured by written examinations on the course content;

.42 with the predictions (ratings) of school instructors as to subsequent suitability as an MR in the fleet.

The best combination of predictors for the composite shipboard criterion proved to be:

1. Fine lathe threading) as reflected by practical
2. Basic milling operations) performance on school work
3. Lathe turning to specifications) samples.

4. Mechanical knowledge) as reflected by the Navy and Guilford-
) Zimmerman Mechanical Knowledge tests.

5. Rated suitability for the MR job) as estimated by instructors
) during Class "A" training.

This produced a multiple R of .62*.

* Estimated by Doolittle method, corrected for shrinkage to .58.

DISCUSSION

There were five basic classes of predictors used in this study. These took the form of scores derived on factors identified in earlier studies performed while the subjects were in basic training. They included:

a. Intellectual aptitude factors
b. Interest factors
c. Measures of academic achievement (written examinations)
d. Measures of academic achievement (school work samples)
e. Instructor ratings of suitability as an MR

Some indication of the complementary nature of the two shipboard criterion measures can be obtained by comparing their correlations with the five classes of predictors. All classes of predictors showed relationships with both the shipboard ratings and the shipboard job performance test in spite of the fact that these two measures did not correlate with each other.

Scores on intellectual aptitude variables correlated more often with the shipboard job performance test scores than with the shipboard ratings. Tests of mechanical knowledge, spatial relations, and perceptual speed all showed significant correlations with the job performance test. In addition, scores on number facility just fell short of producing a significant negative correlation.* By way of contrast, only one intellectual aptitude test, mechanical knowledge, correlated significantly with the ratings assigned by the supervisors aboard ship.

A test of interest in Navy rates developed by Clark failed to correlate significantly with either the shipboard job performance test or the shipboard supervisory ratings taken separately. When the two criteria were combined, however, a significant correlation of .27 with interest in mechanical rates was obtained.

The academic achievement variable was defined by written examinations given in the Class "A" training environment at the end of each two weeks of instruction.

* Negative relationships between mechanical and numerical skills have been noted by other investigators (5, p. 827).

Scores on these tests correlated significantly with shipboard ratings given by supervisors two years later but not with the shipboard job performance test.

Eight group performance factors describing the practical work of the MR trainee in the school environment had been defined in the earlier study (6). Of these, only one, fine lathe threading, correlated significantly with both the shipboard job performance test and the shipboard supervisory rating scale scores. However, scores on factors called hand tool performance and basic milling machine operations correlated significantly with the shipboard job performance test but not with the shipboard ratings. Scores on a factor called lathe-turning-to-specifications correlated significantly with the shipboard ratings but not with the shipboard job performance test. In addition, a score on another school performance factor, basic lathe threading, correlated significantly with the composite shipboard criterion while failing to achieve a significant relationship with either of the criterion measures separately.

Of special interest was the significant correlation found between the ratings given by instructors at the end of each phase of training in the MR School and subsequent performance aboard ship as measured by both the job performance test and the supervisory ratings. [Each instructor was told to rank the trainees in the order in which he would like to acquire them if he were opening up a new machinery repair shop.]

In the school environment, the instructors' ratings had been found to correlate appreciably with both academic achievement as measured by written tests (.63) and with various work sample measures reflecting the practical aspects of MR training (.23 to .57). A question of spuriousness justifiably could be raised concerning these correlations because of the instructors' knowledge of examination grades and intimate acquaintance with the products turned out by the students during training. However, these same instructor rankings correlated .32 two years later with the shipboard job performance test scores and .35 with the ratings given aboard ship

UNCLASSIFIED

by superior petty officers. When the two shipboard criterion measures were combined, the instructor rankings were found to correlate .42 with the resulting scores. This suggests that instructor predictions of subsequent shipboard performance may be a useful intermediate criterion for selection and classification studies, especially if close observation of trainee performance is practiced in the school environment and the ratings are gathered with proper instructions and a well-designed scale.

All of the correlations reported above, while statistically significant, suffer to an unknown degree from the unreliability of both of the shipboard criterion measures. The nature of the shipboard job performance test prohibited its being readministered for purposes of estimating reliability. While it is known that the scoring procedures were highly reliable (.93) and there was evidence of split-half reliability (.87), it is known that testing conditions were far from standard aboard the many ships where data collection took place. It is probable that the performance test suffered from considerable error variance and the ratings from a lack of discrimination among traits. The very low correlation between the two shipboard measures may in part have been due to these limitations.

With such measurement problems, the magnitude of correlations obtained between the various predictors and the unweighted composite criterion is encouraging. Few studies have shown higher relationships between criterion measures and predictors that, on logical grounds, should be related to them. The fact that two years had elapsed between the time when the prediction data were gathered and the shipboard performance measured, adds significance to the results. During this time the subjects had experienced many different job assignments and their opportunity for practice on the various kinds of equipment was a highly uncontrolled variable.

The orientation of this study has been toward the identification of variance in two commonly used, but rarely substantially related, kinds of performance criteria - supervisory ratings and job performance tests. This has been done by

UNCLASSIFIED

examining the relationships between these measures and certain logically related predictors. The evidence presented here suggests that <u>both</u> tests of practical performance and properly executed ratings by superiors very likely are relevant performance criterion measures, and that some combination of the two is superior, on both logical and empirical grounds, to either one alone.

UNCLASSIFIED

Chapter 2

PURPOSE AND METHODS OF THE STUDY

INTRODUCTION

Experience has shown that it is possible to construct discriminating and reliable job performance tests of the practical factors of a Navy rate. Similarly it has been possible to construct rating scales which apparently reflect both the technical and non-technical aspects of shipboard performance. However, the relationships between these two types of performance measures frequently have been found to be so low that one might question the extent of valid variance in either or both kinds of measures.

It is also common for written job knowledge tests to be employed both as criteria of achievement in the training environment and as a basis for promotion in the operating environment. This class of measures frequently fails to correlate substantially with either supervisory ratings or tests of practical performance. However, all three types of measures have shown some relationships with grades obtained in basic training and with a variety of well known aptitude test scores (7, 10).

In an effort to clarify relationships among these various classes of measures under relatively controlled conditions and with more extensive measurement than the shipboard environment usually permits, an earlier study of the performance of 200 Machinery Repairman trainees had been conducted (6). Advantage was taken of the many practical performance tasks required by Class "A" training to make extensive measurements of the practical skills involved and to relate these to written achievement measures, commonly recognized aptitude variables, and rankings by instructors.

Factor analytic studies were conducted of the intercorrelation of these variables. It was found that there was virtually no overlap between the skills

UNCLASSIFIED

measured by practical performance tests and the abilities measured by conventional aptitude tests. It also was observed that the written achievement tests were ineffective for assessing the practical performance skills of the MR trainees.

Ratings of the MR students by the school instructors were found to contain substantial variance in common with a number of the performance measures. However, these ratings also contained reliable variance not associated with practical performance measures and confirmed earlier findings that ratings are factorially complex (1, 13).

Another result of the earlier MR study was that a hypothesized general factor of machinist skills was not found. Rather the practical performance of the trainees seemed best described by a number of relatively narrow common factors. The data also suggested that the factorial nature of machinist performance underwent a change as training progressed. It could be assumed that the modification of skills with advances in training and experience would continue, up to a point, after the trainees were given their regular duty assignments.

Scores on over 100 aptitude, interest, achievement and work sample measures had been accumulated during the MR study. These had been intercorrelated and factor analyzed with the result that 20 factors describing the aptitudes, interests, skills and instructor assessments of the 200 trainees had been identified.

It was felt that this afforded an excellent opportunity for a follow-up study of their shipboard performance. The wide variety of well-defined factors describing the trainees' behavior and aptitudes during training could be related to follow-up measures of the same men, as rated MRs aboard ship, with a possible consequence that some light would be shed upon the nature of the variance in the shipboard performance measures.

UNCLASSIFIED

UNCLASSIFIED

EXPERIMENTAL QUESTIONS

The present investigation was designed to answer the following questions:

1. For a rate such as MR, where the practical performance requirements are clearly defined, how much common variance will be found between scores on a practical job performance test and ratings given by supervising petty officers on various aspects of the total shipboard job?

2. To what extent is the variance in a practical performance test administered aboard ship identifiable from:

   a. practical performance demonstrated during training
   b. aptitude and interest factors
   c. achievement in school as reflected by written tests
   d. predictions (ratings) by instructors of subsequent success as an MR aboard ship?

3. To what extent is the variance in ratings assigned by shipboard supervisors identified by these same predictors?

4. Is it indicated that a composite criterion of shipboard performance should be formed? If so, how much of the reliable variance in the composite criterion measure can be accounted for by these various classes of predictors, singly and in combination?

THE CRITERION MEASURES

The Shipboard Job Performance Test

The Machinery Repairman's responsibilities aboard ship commonly require the fabrication of metal components according to blueprints containing specified tolerances for size, fit, and finish. Construction of these units requires the use of lathes, shapers, drills, hand tools, and milling machines.

It was desired to develop a practical performance test in the form of a blueprint that would call for the construction of an object reasonably like those that might be required of a rated MR and which would sample as many of the skills previously identified in the school environment as practical. With the aid of instructors at the Class "A" MR School, the test blueprint shown in Figure 1 was designed. The skills required in the construction of this product included lathe turning, thread cutting, taper cutting, drilling, milling, cutting an inside taper, and cutting a hemisphere.

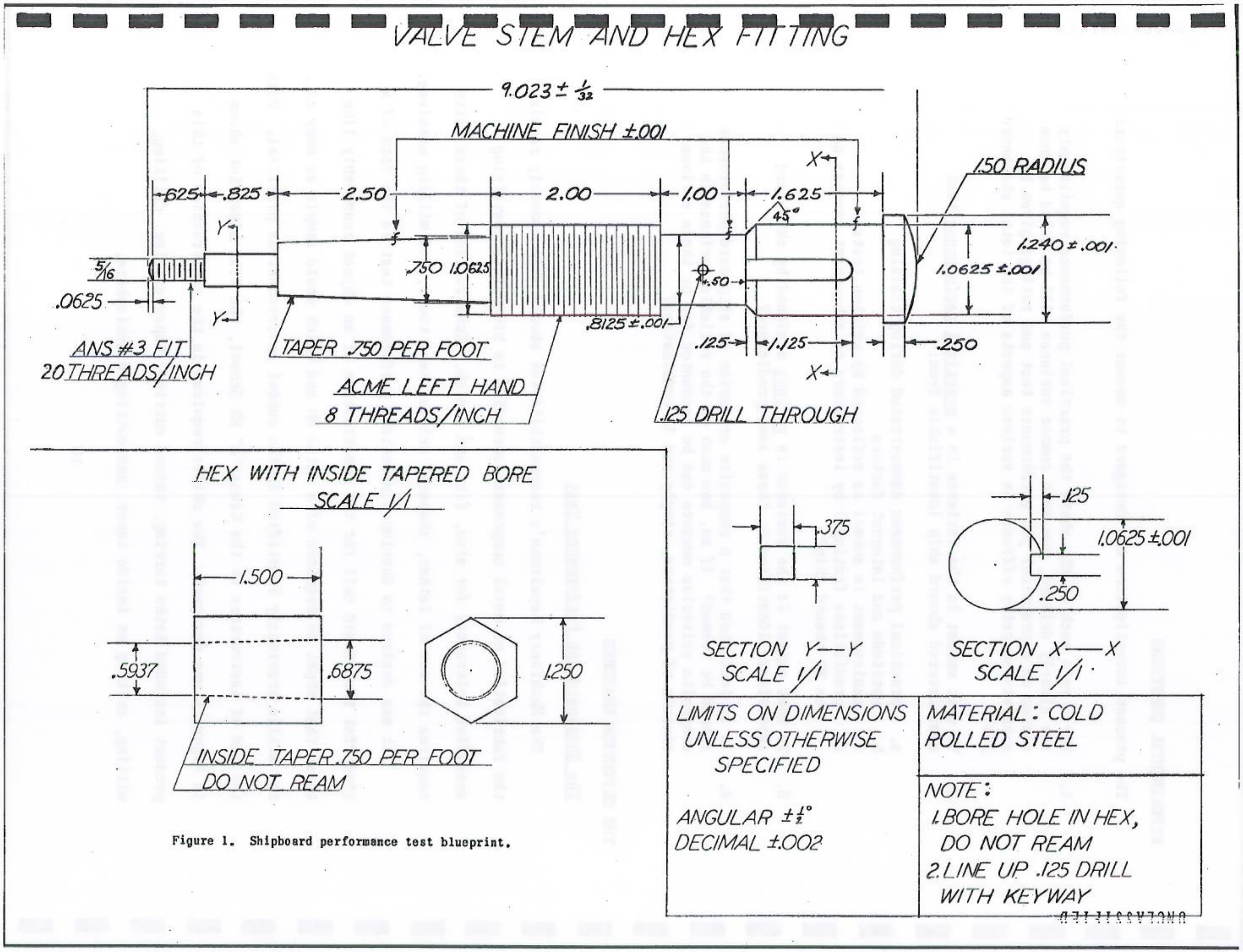

Figure 1. Shipboard performance test blueprint.

UNCLASSIFIED

Administering the Job Performance Test

Since the subjects of this study were aboard ships in widely scattered geographical areas, it was impossible for the research staff personally to supervise all of the performance testing. Wherever possible the ship in question was visited, however, and arrangements were made with the engineering or supervising petty officer for administration of the test.

These supervising personnel were instructed in such matters as the need for strict adherence to the administrative instructions and for insuring the independence of each examinee's efforts.

It was requested that the test product be scheduled as a part of regular ship's work, if at all possible. Estimates by MR School instructors and experience while pre-testing the test had indicated that on the average about 8 hours would be required for the construction of the test product. It became evident early during data collection, however, that the average time required would be closer to 15 hours. For this reason, it proved impossible in most cases to complete the test during one day's work schedule and it was often necessary to schedule the test work during off-duty hours so that it would not interfere with routine shipboard requirements. Both of these conditions helped create non-standard test administration conditions and, to an unknown degree, may have introduced error variance in the performance test scores.

Some of the subjects of the study were on ships in remote areas or were not available in their home ports at the time of the field worker's personal visits. In these cases, arrangements for the performance testing were made by mail. It was recognized that this was contrary to all good test procedure but in view of some of the questions surrounding the use of practical performance tests in the Navy, it was felt that it would be desirable to give this procedure a try.

It should be mentioned that there probably was not much need for detailed administrative instructions with this test. The blueprint, itself, largely stood as a set of instructions. The requirement that the examinee determine the most effective way to go about constructing the product may be considered a part of the test.

Scoring the Job Performance Test

The earlier study of MR performance in the school environment had indicated that many of the performance factors cut across product type and were most readily identified by the kind of operation involved. (See section on predictor variables below.) Consequently it was deemed important to score the performance test so that the various operations involved, if they did reflect different basic skill factors, could be related to the factor scores previously identified in the school environment. For each part of the product, therefore, measurements were taken which would reflect the extent to which diameters, lengths, radii, tapers, and the like conformed to the actual specifications called for in the blueprint. Other parts of the product were scored by assigning judgments to the excellence of finishes. For those parts of the product designed to join with another component, standard gauges were used to assign scores on fit. In all, 33 measures, not including total score, were derived from the test product. These are listed in Table I along with the units of measurement employed, the median score obtained, and Q as a measure of variability.

Because it was desired to minimize as many scores of error variance in the performance measure as possible, all products were scored on all characteristics by two judges independently. These judges employed common scales and definitions insofar as possible in arriving at the score values.

UNCLASSIFIED

UNCLASSIFIED

Table I

Units of Measurement, Averages, Dispersions and Agreement of Judges on Scores Derived from the Shipboard Job Performance Test (N = 64)

	Measure	Units of Measurement	Mdn. Score	Q	Inter-Judge Agreement (rho)
1.	Radius fit	5-pt. scale (rating)	2.9	1.2	.97
2.	" fin.	5-pt. scale	2.0	.9	.96
3.	Flat diameter	.001"	.001	.002	.95
4.	" width	.001"	.010	.011	.95
5.	" fin.	5-pt. scale	2.0	.5	.90
6.	Body length	.005"	.018	.014	.89
7.	" diameter	.001"	.002	.001	.95
8.	" fin.	5-pt. scale	2.0	.8	.88
9.	Bevel fit	" "	1.0	.6	.71
10.	" fin.	" "	2.0	.7	.84
11.	Undercut diam.	.001"	.002	.002	.92
12.	" length	5-pt. scale	1.2	.6	.77
13.	" fin.	" "	2.0	.8	.94
14.	Keyway specs. (l.,d.,w.)	.005"	.034	.018	.95
15.	Pinhole specs. (ctr. angle, align.)	5-pt. scale	1.2	.6	.86
16.	Pinhole diameter	" "	2.0	1.0	.91
17.	Acme thread diam.	.001"	.002	.002	.88
18.	" " length	.005"	.014	.018	.90
19.	" " fin.	5-pt. scale	2.1	.7	.89
20.	Taper length	.005"	.008	.005	.83
21.	" diameter	.002"	.006	.003	.93
22.	" fit	5-pt. scale	3.0	1.0	.99
23.	" fin.	" "	2.0	.9	.91
24.	Square width	.001"	.007	.005	.98
25.	" length	.005"	.017	.020	.92
26.	" alignment	5-pt. scale	1.6	1.4	.93
27.	ANS thread length	.005"	.018	.020	.70
28.	" " diam.	.001"	.004	.003	.92
29.	" " fin.	5-pt. scale	2.0	.6	.82
30.	Overall length	" "	1.8	.7	.96
31.	Hex length	.003"	.008	.008	.96
32.	" diagonal	.010"	.044	.040	.98
33.	Inside taper fit	5-pt. scale	2.9	1.0	.94
				Median	.92
34.	Total score				.98

The accuracy of their scoring procedures is reflected in the last column of Table I where the extent of their agreement on the rank order of the products for each characteristic is reflected by rho coefficients. These coefficients ranged from .70 to .99 with a median value of .92. Agreement on the score for total product was .98.

It was possible to utilize three kinds of scores from the test product.

1. A score for each of the 33 individual measurements.
2. Scores for certain combinations of measures that produced sub-totals thought to be related to the performance factors observed in the school environment.
3. Total score for the entire test product.

The Shipboard Performance Rating Scale

The shipboard performance rating scale followed a design similar to that employed in previous studies of shipboard criteria which had been found to provide discriminating and reliable results (8). The scale was in graphic form and required that the supervising petty officer rate his men by locating each one with respect to 25 hypothetical MR's that were lined up such that they ranged from the best to the poorest MR in the Navy (See Figure 2). This was done for each of 15 traits of performance in turn. The instructions emphasized that the rater should compare each man rated with respect to all machinery repairmen that he had known of the same pay grade. Verbal descriptions of hypothetical cases were printed on the rating form at both extremes and at the mid-point of each scale to serve as anchoring definitions for the raters.

Since the number of men belonging to the study sample found aboard any one ship ranged from 1 to 8, the opportunity to make man-to-man comparisons differed for different raters. In those cases where only one examinee was

UNCLASSIFIED

UNCLASSIFIED

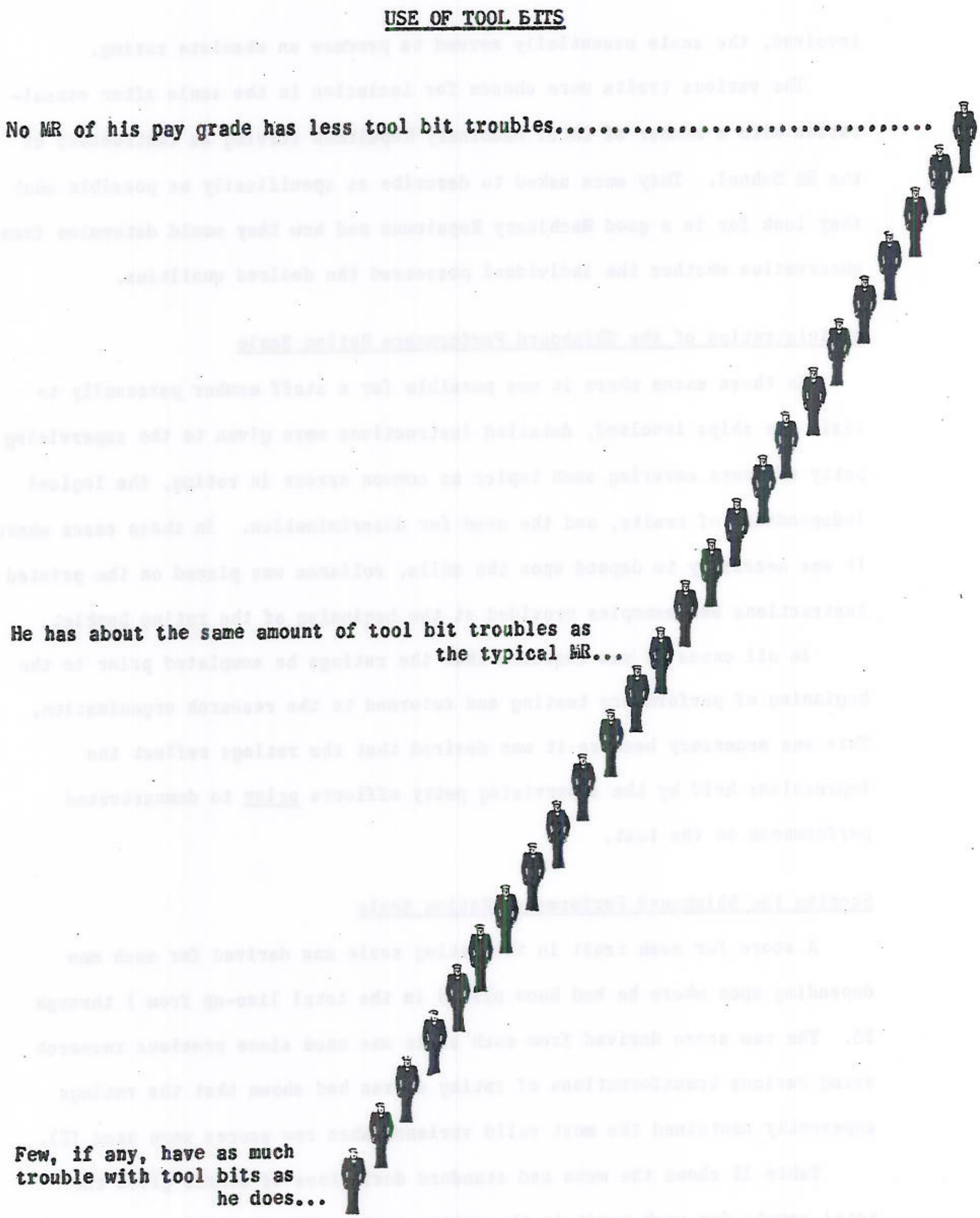

Figure 2. Sample page from shipboard performance rating scale.

involved, the scale essentially served to produce an absolute rating.

The various traits were chosen for inclusion in the scale after consultation with a number of Chief Machinery Repairmen serving as instructors at the MR School. They were asked to describe as specifically as possible what they look for in a good Machinery Repairman and how they would determine from observation whether the individual possessed the desired qualities.

Administration of the Shipboard Performance Rating Scale

In those cases where it was possible for a staff member personally to visit the ships involved, detailed instructions were given to the supervising petty officers covering such topics as common errors in rating, the logical independence of traits, and the need for discrimination. In those cases where it was necessary to depend upon the mails, reliance was placed on the printed instructions and examples provided at the beginning of the rating booklet.

In all cases it was required that the ratings be completed prior to the beginning of performance testing and returned to the research organization. This was necessary because it was desired that the ratings reflect the impressions held by the supervising petty officers prior to demonstrated performance on the test.

Scoring the Shipboard Performance Rating Scale

A score for each trait in the rating scale was derived for each man depending upon where he had been placed in the total line-up from 1 through 25. The raw score derived from each scale was used since previous research using various transformations of rating scores had shown that the ratings apparently contained the most valid variance when raw scores were used (7).

Table II shows the mean and standard deviations of scores given the total sample for each trait in the rating scale. The physical mid-point of

UNCLASSIFIED

UNCLASSIFIED

the rating scale would be represented by a score of 13.0. It can be seen from Table II that there was some leniency effect, the mean ratings assigned to the various traits ranging from 14.4 to 17.5. This effect was slight compared to that obtained with many rating scales, however, and the discrimination represented by an average standard deviation of better than 4.0 suggested that the raters were reasonably realistic and discriminating in their appraisals.

Table II

Means and Standard Deviations of Shipboard Rating Scale Trait Scores

	M	σ
1. Pride in Work	17.5	4.0
2. Response to Authority	17.0	4.7
3. Persistence with a Job	17.6	4.3
4. Acceptance of Responsibility for Mistakes	17.0	4.7
5. Lack of Need for Supervision	17.0	4.3
6. Planning a Job	16.6	4.5
7. Use of Tool Bits	16.4	4.1
8. Ability to Produce a High-Grade Finish	15.8	3.0
9. Working to Specifications	16.6	4.5
10. Skill in Reading Blueprints	15.9	4.2
11. Care of Tools and Machinery	16.6	4.5
12. Use of a Lathe	16.9	4.0
13. Use of a Milling Machine	15.2	4.8
14. Use of a Shaper	14.4	4.7
15. Job Knowledge	17.0	3.8

THE PREDICTOR VARIABLES

It will be recalled that the predictor data used in this study had been generated two years earlier in the Class "A" training environment. In an attempt to answer questions concerning the nature of variance in practical performance measures, a factor analysis had been conducted of the intercorrelations of aptitude and interest test scores, measures of practical performance required by the school curriculum, formal written achievement test scores, and instructor appraisals of the potential worth of each trainee as an MR. This had resulted in the identification of five broad classes of factors as shown in Table III:

Table III

Predictor Factors Together with Test Variables Leading Highest on Each Factor

I Aptitude Factors

1) Verbal Comprehension
   a) Guilford-Zimmerman, Verbal Comprehension
   b) Navy GCT
2) General Reasoning
   a) Guilford-Zimmerman, General Reasoning
   b) Navy Arithmetic Reasoning
3) Number Facility
   a) N-3 Subtraction and Multiplication
   b) N-1 Addition
4) Mechanical Knowledge
   a) Guilford-Zimmerman, Mechanical Knowledge
   b) SRA Mechanical Comprehension
5) Spatial Relations
   a) Vz-2 Punched Holes
   b) S-1 Cards
6) Associative Memory
   a) PMA First-Names
   b) PMA Word-Number
7) Perceptual Speed
   a) Guilford-Zimmerman, Perceptual Speed
   b) FACT 1A, Inspection
8) Motor Control
   a) Ai-2, Tracing, Easy
   b) Ai-3, Tracing, Difficult

II Interest Factors

9) Interest in Mechanical Rates
   a) Engineman Interest
   b) Gunner's Mate Interest
10) Interest in Mechanical Activity
   a) Interest in Mechanical Activity
   b) Interest in Manual Activity
   c) Interest in Precision

III Work Sample Factors

11) Lathe Threading, Basic
   a) C-Clamp Screw-Fit
   b) C-Clamp Screw-Specifications
12) Lathe Threading, Fine
   a) Parallel Clamp Screw-Fit
   b) Parallel Clamp Screw-Specifications
13) Hand Tool Performance
   a) Parallel Clamp-Finish
   b) C-Clamp-Finish
14) Lathe Precision Turning
   a) Mandrel and Fits-Specifications
   b) Mandrel and Fits-Fit
15) Lathe Turning to Specifications
   a) Morse Center-Specifications
   b) Morse Gauge-Specifications
   c) Gear Blank-Specifications
16) Lathe Turning to Finish
   a) Morse Center-Finish
   b) Morse Gauge-Finish
17) Milling Machine, Basic
   a) Male Blocks-Specifications
   b) Female Blocks-Specifications
18) Milling Machine, Complex Indexing
   a) Square
   b) Hex Nut

IV Academic Achievement Factor

19) MR School Achievement (Written Tests)
   a) MR School Examination (Phase VI)
   b) " " " (Phase I)
   c) " " " (Phase III)
   d) Purdue Milling Machine

V Instructor Assessment Factor

20) Rated Suitability for the Job
   a) Supervisor Rank (Phase II)
   b) " " (Phase IV)
   c) " " (Phase V)
   d) " " (Phase VI)

The aptitude and interest factors were defined by well known published tests whose factor content had been identified in previous studies. The work sample factors, of course, had been discovered for the first time in the study of trainee performance. It was not known to what extent these factors would be revealed in the subsequent shipboard performance of rated MR's. The same was true of the academic achievement factor which had been defined mainly by the school's regular phase-end written examinations.

It was of interest also to learn whether the fifth class of predictors defined by appraisals of school instructors would be predictive of various aspects of subsequent shipboard performance. In the school environment they had been related substantially to both the quality of practical work performed and to formal achievement as measured by written tests.

A score on each of the 20 predictor factors was derived for each subject according to Catell's method for estimating factor endowments by summing the scores of individuals on the high loading tests for each factor. The high loading tests used for this purpose also are identified in Table III. (More complete descriptions of the individual tests has been given in an earlier report (6).) A test was not included for estimating factor scores if it showed substantial loadings on more than one common factor. In addition an attempt was made to use only those tests whose factor saturations were not markedly dissimilar. Equal weight was given to each test. The summed scores for each factor were then transformed to centiles to provide a common median and range for the several distributions.

THE SUBJECTS

64 rated Machinery Repairmen who had previously been tested as students at the Class "A" Machinery Repair School, Naval Training Center, San Diego, served as subjects for the study. Since graduation, these men had been serving their enlistments aboard a variety of Naval vessels including light and heavy repair ships,

UNCLASSIFIED

aircraft carriers, seaplane tenders, destroyers, destroyer escorts, and floating dry docks. Each man had been on active duty a minimum of two years since being graduated from the school.

During the data collection phase of this follow-up study, it was possible to locate approximately 120 of the original 200 Machinery Repairman students. Field trips were made to Long Beach, San Diego, San Francisco, Norfolk, New York, Philadelphia, Newport, Providence, and Boston in an attempt to make personal contact with the supervisors of as many of these men as possible. As previously indicated, where this was not possible an attempt was made to set up the test and rating procedures by mail.

Discharges, transfers, disciplinary actions, and prior commitments of various kinds, together with incomplete data on some cases, resulted in a reduction of this potential sample from 120 to 64. Although this number was smaller than was desired, it was felt that the wealth of data available on each subject fully justified the follow-up study.

A comparison between the scores of the predictor variables of the 64 subjects comprising the follow-up sample and those of the original group of 200 tested in the training environment indicated that the follow-up sample was unbiased insofar as basic aptitudes, interests and skills were concerned. Using the 46 tests employed to derive factor scores, it was found that the median scores of the follow-up sample were significantly greater than .50 (the median for all centile scores) for only two tests. It was evident from inspection of the score distributions that differences between medians of the follow-up sample and its parent group were within the standard error of measurement.

Of the 64 men comprising the follow-up sample, 27 were second class MR's, 23 were third class MR's, 4 were strikers, and 10 did not report at the time of the study. To some extent this represented differential shipboard experience although as pointed out previously all men had had a minimum of 2 years of sea duty.

UNCLASSIFIED

STATISTICAL PROCEDURES

The study took the form of a straightforward correlational and multiple regression analysis. This required the determination of:

1. the most suitable scores to use for each of the two shipboard performance criterion measures.
2. the validity of each predictor or class of predictors for the two criterion measures.
3. the intercorrelations of the predictors.
4. the best combination of predictors for each criterion variable.
5. the best composite criterion.
6. the best combination of predictors for the composite criterion.

All necessary intercorrelation matrices were computed on the SWAC at UCLA by Dr. Andrew L. Comrey. All other statistical work was accomplished by calculator, multiple R's being estimated by the Gingerelli approximation or the Doolittle method as seemed appropriate.

UNCLASSIFIED

Chapter 3

RESULTS I: INTERCORRELATIONS AND RELIABILITIES OF MEASURES

CRITERIA

Intercorrelations of Shipboard Job Performance Test Scores

It will be recalled that, in developing the performance test for shipboard use, an attempt was made to encompass as many of the practical activities of the Machinery Repairmen as possible in a single test product. An attempt also was made to reflect those performance factors that previously had been identified from the work sample measures in the training environment.

Since time limitations made it impractical to have each subject produce more than one product, any attempt to relate scores on the performance factors found in the school environment to performance on the shipboard test required that scores on the latter be broken down to reflect the various kinds of skills involved. 33 measures had been made in scoring the shipboard product. It was felt that these could be logically combined into certain subtotal scores that might logically be related to the performance skills identified in the school environment.

It was of interest first to compute the intercorrelations between these subtotal scores to determine whether relatively independent factors of performance were to be found in the behavior of skilled repairmen as had been found with trainees. In Table IV the intercorrelations of scores on the various suboperations of the shipboard performance test are presented. They do not differ greatly in magnitude from intercorrelations reported with other types of performance test batteries where a broad sampling of the practical performance of a rating has been attempted (11, 12). The general magnitude of correlation probably is no greater than would have been expected considering that the factor analysis of work sample measures in the school environment had required

UNCLASSIFIED

Table IV

Intercorrelations of Shipboard Performance Test Scores for Various Operations

	1	2	3	4	5	6	7	8
1. Lathe Threading, Basic								
2. Lathe Turning to Specifications	37							
3. Lathe Turning, Basic	34	38						
4. Drilling	16	26	28					
5. Milling Operations	14	32	37	17				
6. Lathe Turning to Finish	48	27	35	22	33			
7. Acme Thread Specifications	02	23	42	03	37	33		
8. Radius Specifications	03	16	23	17	57	16	18	

an oblique solution. In any event the correlations were such that it seemed reasonable to proceed with a test of the hypothesis that the several skills identified from the work sample measures in school would be related to scores made on the various sub-operations on the shipboard job performance test.

Before this hypothesis could be tested, however, the general questions of the reliability of performance on these several sub-operations had to be considered. It was entirely possible, of course, that the relatively low intercorrelations obtained might have reflected little more than unreliability among the measures.

Reliability of the Shipboard Job Performance Test

Lengthy administration time and the fact that the job performance test required the utilization of shipboard machinery that was heavily programmed made it impossible to obtain retest reliability measures of job performance for this study. A total of 33 measures was made on each test product, however, making it possible to obtain some split-half estimates. It must be admitted that the half-tests were insufficiently long to place much confidence in these estimates and questions also could be raised concerning their comparability in terms of difficulty. Nevertheless it was felt that some evidence of reliability would be better than none at all. The split-half estimates, corrected for double length, are shown in Table V. The various measures that make up the two half-scores are shown in each case.

Table V

Split-Half Reliabilities for Various Sub-Operations of the Shipboard Job Performance Test

Operation	r_{tt}
<u>Lathe Threading, Basic</u>	.26
ANS thread length	
ANS thread diameter	
ANS thread finish	
Bevel fit	
<u>Lathe Turning to Specifications</u>	.51
Taper length	
Taper diameters	
Taper fit	
Hex fit (inside bore)	
<u>Lathe Turning, Basic</u>	.65
Body length	
Body diameter	
Flat diameter	
Flat width	
Undercut diameter	
Undercut length	
<u>Drilling</u>	.36
Pinhole alignment	
Pinhole diameter	
<u>Milling</u>	.41
Hex nut diagonal	
Hex nut length	
Keyway depth	
Keyway width	
Keyway length	
Square width	
Square length	
<u>Lathe Turning to Finish</u>	.81
Taper finish	
Body finish	
Undercut finish	
Flat finish	
ANS thread finish	
Bevel finish	
<u>Acme Thread Specifications</u>	.19
Acme thread diameter	
Acme thread length	
<u>Radius Specifications</u>	.95
Radius fit	
Radius finish	

<u>Decision to Use Only the Total Job Performance Test Score</u>

The intercorrelation analysis and the split-half reliability computations had indicated that it was at least feasible to go ahead with a study of relationships between the various predictor variables and the shipboard job performance test scores broken down by type of operation. The results of

UNCLASSIFIED

this correlational analysis are shown in Table VI. In this matrix a coefficient of .25 is required for significance at the .05 level of confidence.

Table VI

Correlations Between Scores on Sub-Operations of Shipboard Performance Test and Similar Work Operations During Class "A" Training

Class "A" School Training Operation	Shipboard Performance Operation								
	Lathe Threading Basic	Lathe Turning to Spec	Lathe Turning Basic	Drill-ing	Mill-ing	Lathe Turning to Fin.	Acme Thread Specs.	Radius Specs.	Total Product Score
Lathe Threading, Basic	06	18	01	26	13	13	23	15	21
Lathe Threading, Fine	15	22	20	40	09	26	24	15	34
Hand Tool Performance	23	25	09	10	12	19	26	21	30
Lathe Precision Turning	14	15	19	-08	10	07	21	04	16
Lathe Turning to Specs.	-04	-11	10	30	05	02	08	-03	04
Lathe Turning to Finish	14	17	-01	15	-12	23	-01	-18	08
Milling Machine, Basic	29	26	34	40	20	18	18	02	37
Milling Machine, Complex Indexing	29	03	16	19	-06	06	09	-05	12

While many of the correlations in Table VI are significant, it is evident that there are no systematic patterns of relationships between the work sample scores on practical operations in MR School and the apparently similar types of scores produced by the operational breakdown of the shipboard job performance test. For example, while Basic Lathe Threading as measured aboard ship shows two significant correlations with factor scores from practical performance in the school environment, it failed to correlate significantly with the factor identified as Basic Lathe Threading in the school environment.

This lack of systematic and logical relationships can be observed

UNCLASSIFIED

UNCLASSIFIED

throughout the table of correlations. The results of this analysis, and from the fact that the breakdown of the shipboard job performance test into sub-operations had to be accomplished rationally rather than empirically, led to the conclusion that only the <u>total</u> score from the shipboard job performance test should be used in subsequent analyses. The marginal reliabilities of the sub-operation scores strengthened this conviction, and, at the same time, raised the question of the reliability of the total performance test score.*

A split-half reliability estimate for the total performance test score was made by assigning the 33 separate scores derived from the product into two halves such that approximately the same number of scores for lengths, diameters, fits and finish was represented in each half. In addition an attempt was made to represent each basic type of operation in the two halves (see Table VII). When this was done the resulting reliability (corrected) for the shipboard performance test total score was .87.

While it is recognized that the assumptions for split-half reliability estimates can hardly be justified with a test that is known to be factorially complex, it is again felt that this estimate is better than nothing. Although factorial complexity might act to produce a low split-half estimate, the likelihood of corrected errors in the two arbitrary halves of a performance test of this kind makes it desirable to regard this value as an upper bound of the estimated reliability.

---

* A similar analysis relating the 33 separate scores derived from the various parts of the test product to the various work sample scores made in Class "A" training produced almost no significant correlations. This made the reliability of these single scores highly suspect. The greater number of significant correlations appearing in the analysis reported above may have been a function of improved reliability as more items were included in the job performance test score. This line of reasoning would also support the conclusion that only <u>total</u> score on the job performance test should be used for future analyses.

UNCLASSIFIED

Table VII

Division of Scores Into Halves for Reliability Estimate of Shipboard Job Performance Test Score

A	B
ANS thread length ANS thread finish	ANS thread diameter
Taper length Taper fit	Taper diameter Taper finish
Body length Body finish	Body diameter
Flat diameter Flat finish	Flat width
Undercut diameter	Undercut length Undercut finish
Pinhole alignment	Pinhole diameter
Hex nut diagonal	Hex nut fit (inside bore) Hex nut length
Keyway depth	Keyway width Keyway length
Square length	Square width
Acme thread finish Acme thread diameter	Acme thread length
Bevel finish	Bevel fit
Radius fit	Radius finish

r_{tt} = .87 (corrected for double length)

<u>Intercorrelations of the Shipboard Rating Scale Trait Scores</u>

In Table VIII the intercorrelations of trait scores from the shipboard rating scale are presented. Inspection readily reveals the frequently found high intercorrelations among rated traits. Only on <u>Use of a Shaper</u>, a piece of equipment not available in many of the shipboard settings, did the correlations systematically drop below the .50 to .80 range.

Table VIII

Intercorrelations Among Shipboard Rating Scale Trait Scores

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Pride in Work															
2	Response to Authority	69														
3	Persistence with a Job	74	67													
4	Acceptance of Responsibility for Mistakes	63	55	74												
5	Need for Supervision	71	62	87	76											
6	Planning a Job	64	51	68	58	69										
7	Use of Tool Bits	65	45	67	50	69	63									
8	Ability to Produce High-Grade Finish	69	55	70	55	72	64	81								
9	Working to Specifications	77	61	77	64	77	72	81	88							
10	Skill in Reading Blueprints	66	54	75	74	77	69	68	71	75						
11	Care of Tools and Machinery	70	70	72	70	73	56	56	65	65	62					
12	Use of a Lathe	65	54	76	63	79	65	80	84	84	76	69				
13	Use of a Milling Machine	54	54	62	43	61	52	58	67	62	54	56	65			
14	Use of a Shaper	37	32	43	37	36	34	49	37	39	34	42	32	49		
15	Job Knowledge	72	65	75	62	71	56	62	59	71	58	58	66	46	39	

In spite of this general magnitude of relationships, more than one factor of rated behavior may be represented in this matrix. Because it did not seem warranted, from the point of view of investment in either time or funds, the factor analysis which would have answered this question was not performed. Instead it was decided that differential patterns of correlations would be looked for in the already computed relationships between scores on the various rating scale traits and scores on the several sub-operations of the shipboard job performance test.

It was argued that if different factors of performance were represented in the rating scale scores, differential patterns of correlations should have been obtained with the several scores derived from the shipboard job performance test. The failure of this hypothesis to be verified is revealed in Table IX. With the exception of the correlations between the score on drilling

Table IX

Correlations Between Scores on the Shipboard Rating Scale Traits and Scores on the Several Operations of the Shipboard Job Performance Test

Rating Scale Traits	Operations on Job Performance Test								
	Lathe Threading Basic	Lathe Turning to Spec	Lathe Turning Basic	Drilling	Milling	Lathe Turning to Fin.	Acme Thread Specs.	Radius Specs.	Total Score
1 Pride in Work	-02	07	06	27	-09	11	07	04	07
2 Response to Authority	-06	02	-04	13	-19	-03	-03	-05	-08
3 Persistence with a Job	-19	14	-11	31	-06	-01	-04	03	00
4 Acceptance of Responsibility for Mistakes	-17	14	-25	12	00	-02	03	05	-02
5 Need for Supervision	-15	18	-19	30	-15	-06	-22	-02	-08
6 Planning a Job	-06	20	05	18	07	21	12	06	19
7 Use of Tool Bits	00	16	12	24	-08	-02	-07	-05	06
8 Ability to Produce a High-Grade Finish	-15	14	09	38	-07	04	-03	07	08
9 Working to Specifications	-11	09	11	30	-07	03	-02	-03	05
10 Skill in Reading Blueprints	-11	03	-14	17	-13	03	-07	-02	-03
11 Care of Tools & Machinery	-14	08	-05	21	01	-14	-10	09	-03
12 Use of a Lathe	-12	14	04	22	-09	-00	-09	-10	02
13 Use of a Milling Machine	-17	05	04	32	-09	01	-06	00	01
14 Use of a Shaper	-03	14	11	34	07	-02	08	13	13
15 Job Knowledge	-11	21	-04	11	-10	-05	04	06	01

UNCLASSIFIED

operations and the several rating scale traits, few significant relationships of any kind will be observed.

Particularly noteworthy was the lack of relationship between the several rating scale trait scores and total score on the job performance product. These correlations ranged from -.08 to .19, none of them being significantly different from zero.

Reliability of the Shipboard Rating Scale Total Score

This result raised again the question of the reliability of the criterion measures, this time of the scores on the rating scale traits. Because there was a general lack of diagnostic quality in the rating scale scores, as reflected by the uniformly high intercorrelations, it was decided that in subsequent analyses a simple average of the 15 rating scores would be used in an effort to insure reliability in the rating measure.

An estimate of internal consistency of the rating scores was made by randomly dividing the scales into two groups, summing the scores for each group of traits, and correlating the two halves. This produced a coefficient, corrected for double length, of .97 indicating that whatever the rating scale measured, it measured with a high degree of consistency.

PREDICTORS

Intercorrelations of Predictors

In Table X the intercorrelations among the 20 predictor variables derived from measures administered during the Class "A" School training of the subjects are presented. For the most part these intercorrelations were relatively low, which was to be expected since the predictor scores represented endowments on factors that previously had been found to be largely independent.

Only a few correlations above .40 will be noted. These include a correlation between Mechanical Knowledge and Spatial Relations of .47,

UNCLASSIFIED

Table X

Intercorrelations among Predictor Scores
(N = 64)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	Verbal Comprehension																				
2	General Reasoning	20																			
3	Number Facility	36	39																		
4	Mechanical Knowledge	43	22	-08																	
5	Spatial Relations	23	45	08	47																
6	Associative Memory	18	22	33	08	-18															
7	Perceptual Speed	16	01	13	38	18	02														
8	Motor Control	18	04	29	05	09	21	15													
9	Interest in Mechanical Rates	-11	-01	-07	39	14	19	04	-02												
10	Interest in Mechanical Activity	-13	00	-02	11	11	-11	-12	04	33											
11	Lathe Threading, Basic	06	15	06	21	13	08	21	09	22	-04										
12	Lathe Threading, Fine	-10	12	-16	24	19	-09	21	02	10	-20	54									
13	Hand Tool Performance	17	15	-08	29	17	08	24	03	09	-22	40	41								
14	Lathe Precision Turning	14	-02	-02	03	-01	08	20	14	-04	-19	04	17	26							
15	Lathe Turning to Specifications	-05	-06	-14	22	05	-10	04	-05	23	02	15	36	23	23						
16	Lathe Turning to Finish	-04	03	-16	04	-05	-10	02	-10	12	09	20	24	37	26	24					
17	Milling Machine, Basic	11	20	-03	24	25	13	-13	-17	17	05	20	28	13	05	25	03				
18	Milling Machine, Complex	18	20	08	24	12	15	08	-16	16	-02	28	14	14	08	05	26	27			
19	MR School Achievement	52	37	31	43	40	21	14	10	17	01	37	18	35	20	17	11	19	32		
20	Rated Suitability for the Job	41	13	09	38	30	02	19	01	16	-05	45	30	45	24	32	28	23	57	63	

a correlation that is often noted; a correlation of .45 between General Reasoning and Spatial Relations, also a common finding; and a correlation of .43 between Verbal Comprehension and Mechanical Knowledge, a not frequently encountered finding. This last relationship has been interpreted to mean that restricted verbal skills in this sample of personnel were responsible for inducing a loading in the mechanical knowledge tests on the verbal factor.

There are a number of other moderate correlations in this matrix that are of general interest. A correlation of .41 between Verbal Comprehension and the prediction by MR School instructors of the Suitability of each trainee for performing the work of an MR perhaps reveals an important degree of reliance on intellectual considerations in rating a man's capabilities for practical performance. The correlation of .52 between Verbal Comprehension and scores on the MR Academic Achievement examinations suggests a reason for the relationship observed. In this sample, at least, achievement on the written examinations was dependent upon verbal comprehension to a much higher degree than might be desired. The school instructors, having knowledge of academic grades, thus produced ratings that were in turn loaded on the verbal comprehension factor.

The MR School achievement test scores correlated .63 with the instructor predictions of Suitability as an MR. In addition they correlated .40 or higher with three aptitude test variables, Mechanical Knowledge, Spatial Relations, and Verbal Comprehension.

Only two other predictor variables frequently showed relationships with other predictors above .40. Basic Lathe Threading correlated .54 with Fine Lathe Threading, .40 with Hand Tool Performance and .45 with instructor predictions of Suitability as an MR. Hand Tool Performance, in addition to its correlation with Basic Lathe Threading, correlated .41 with Fine Lathe

UNCLASSIFIED

Threading and .45 with rated Suitability for the MR job.

The instructors' ratings or predictions of eventual Suitability as an MR proved to be the most complex factorially of all the predictors. As previously mentioned, they correlated .41 with Verbal Comprehension, .45 with Basic Lathe Threading, .45 with Hand Tool Performance, .57 with Complex Milling Machine operations, and .63 with the MR written achievement examinations. In addition, these ratings correlated significantly with tests of Mechanical Knowledge, Spatial Relations, Fine Lathe Threading, and Lathe Turning to Specifications.

These indications of the complex factorial composition of the school instructors' ratings are of considerable interest in themselves. The substantial correlations with performance factors may have resulted from the fact that the instructors' responsibilities required them to be closely familiar with the work samples produced by the trainees. In addition they had access to information regarding the trainee's achievement as reflected by the school's written subject matter tests. The high validity of their ratings for the various school criteria might therefore be regarded as a result of atypical (if not spurious) opportunities to have knowledge of trainee performance. Subsequent analyses showed, however, that these same instructor ratings bore important relationships to performance aboard ship two years later, suggesting that, whatever the reason, they reflected impressions that were important for eventual success.

UNCLASSIFIED

UNCLASSIFIED

Chapter 4

RESULTS II: RELATIONSHIPS BETWEEN THE SHIPBOARD CRITERION MEASURES AND VARIOUS PREDICTORS

The analyses reported in Chapter 3 had indicated that: (1) both shipboard criterion measures were characterized by a high degree of internal consistency; (2) the criterion measures were essentially independent of each other; (3) the several predictors were largely independent of each other and therefore might shed some light on the differential nature of the shipboard criterion measures. Correlational analyses were conducted, therefore, between the various classes of predictors and the two shipboard performance measures, singly and in combination.

CORRELATIONS OF THE SHIPBOARD CRITERIA WITH APTITUDE FACTORS

Job Performance Test

Total score on the shipboard job performance test correlated significantly with three of the eight aptitude factors as shown in Table XI. The three tests showing significant relationships with the job performance test, Mechanical Knowledge, Spatial Relations and Perceptual Speed, appear quite logically related to performance requiring mechanical skills. Together they produced a multiple R of .39 with total score on the performance test.

Table XI

Correlation of Shipboard Criterion Measures with Aptitude Factors (N = 64)

Aptitude	Job Performance Test	Average Rating Score
1 Verbal Comprehension	.08	.18
2 General Reasoning	-.12	.01
3 Number Facility	-.24	.13
4 Mechanical Knowledge	.36**	.29*
5 Spatial Relations	.32*	.17
6 Associative Memory	-.03	.11
7 Perceptual Speed	.32*	.14
8 Motor Control	-.18	.12

* significant at 5% level
** significant at 1% level

It is noteworthy that Number Facility just missed correlating significantly in a negative direction with the job performance test. Considered from the viewpoint of the job to be performed, which requires a certain amount of computational skill in setting up the work, this is difficult to comprehend. On the other hand, other investigators have noted slight negative correlations between measures of mechanical and numerical aptitude (5).

Supervisory Ratings

Only one aptitude variable, Mechanical Knowledge, correlated significantly with the ratings assigned by supervising petty officers aboard ship. It is of interest that the tests of Mechanical Knowledge used showed validity for both criterion measures in spite of the fact that the subjects had been pre-selected on the Navy Mechanical test. It is also of interest that the ratings of supervisors aboard ship apparently were less dependent upon variance of an intellectual kind than those assigned by instructors in the MR School two years earlier (see Chapter 3).

CORRELATIONS OF THE SHIPBOARD CRITERIA WITH CLASS "A" SCHOOL WORK SAMPLE FACTORS

Job Performance Test

In Table XII it will be observed that scores on the shipboard job performance test correlated significantly with three of the eight performance factors previously identified in the Class "A" School environment. Taken together, Basic Milling Machine operations, Fine Lathe Threading, and Hand Tool Performance, as measured in the school environment, predicted performance on the shipboard job performance test with a multiple R of .44. As suggested by the discussion in Chapter 3, it is not possible from the data generated by this study to associate specific performance operations identified in the school environment with specific operations on the shipboard performance test. However, since these significant correlations cut across all major operations commonly engaged in by MR's aboard ship they are regarded as particularly significant from a logical point of view.

UNCLASSIFIED

UNCLASSIFIED

Table XII

Correlation of Shipboard Criterion
Measures with School Work Sample Factors
(N = 64)

	Work Sample Factor	Job Performance Test	Average Rating Score
11	Lathe Threading, Basic	.21	.20
12	Lathe Threading, Fine	.34**	.32*
13	Hand Tool Performance	.30*	.06
14	Lathe Precision Turning	.16	.20
15	Lathe Turning to Specifications	.04	.40**
16	Lathe Turning to Finish	.08	.09
17	Milling Machine, Basic	.37**	.23
18	Milling Machine, Complex	.12	.11

* significant at 5% level
** significant at 1% level

<u>Supervisory Ratings</u>

The shipboard performance rating scale produced two significant correlations with the performance factors identified in the training environment plus three near-misses. A combination of scores on Lathe Turning to Specifications and Fine Lathe Threading produced a multiple R of .<u>44</u> with the shipboard rating scale total score.

The most striking difference between the correlation of the work sample factors with the two shipboard criterion measures occurred with Lathe Turning to Specifications. The variance associated with this factor was present to a very significant degree in the shipboard ratings but not in the job performance test. Whether this difference would stand up under repeated observations cannot be stated with confidence. It is conceivable, though, that adherence to specifications could be regarded as an extremely important trait by the shipboard supervisors while producing very little variance in the job performance test score.

A reverse argument could be used to explain the difference in common variance shared by the shipboard job performance test and the shipboard ratings with scores on the Hand Tool Performance factor. It is quite conceivable that skill with hand tools among a group of rated petty officers could be taken for granted by their supervisors. In any event, differential skill with hand tools might not be as readily observable as differences in ability associated with the operation of machines.

CORRELATIONS OF THE SHIPBOARD CRITERIA WITH INTEREST, ACHIEVEMENT, AND INSTRUCTOR ASSESSMENT FACTORS

This group of predictors included factors defined by the written achievement examinations covering the course content for various phases of instruction at the MR School; two tests of interest, one developed by Clark for Navy rates, and one by Guilford, et al, for Air Force personnel; and the ratings by school instructors of each man on his suitability as a future machinery repairman.

[Each instructor had been asked to rank his students in the order in which he would like to acquire them if he were opening up a new machine shop. This instruction was designed deliberately to foster a consideration of <u>all</u> of the desirable characteristics of men performing in a machine shop rather than simply the more narrow aspects of the job as reflected by the mechanical skills involved.]

Job Performance Test

In Table XIII it can be seen that of this heterogeneous group of predictors only the instructors' ratings of suitability, made two years earlier, was significantly predictive of performance aboard ship as measured by the job performance test scores. Neither the interest factors nor the formal school achievement factor correlated significantly with practical performance as measured.

Supervisory Ratings

Two of this group of predictors correlated very significantly with ratings given by shipboard supervising petty officers two years later. The

UNCLASSIFIED

ratings of school instructors agreed significantly with the shipboard ratings as did scores on the formal achievement tests given in the MR School. Again, neither of the interest factors correlated significantly with the shipboard rating scores.

Table XIII

Correlation of Shipboard Criterion Measures with Interest, Achievement and Instructor Rating Factors (N = 64)

Predictor		Job Performance Test	Average Rating Score
9	Interest in Mechanical Rates	.23	.17
10	Interest in Mechanical Activity	-.10	.14
19	MR School Achievement Examination	.19	.35**
20	Rated Suitability for the MR Job	.32*	.35**

* significant at 5% level
** significant at 1% level

MAXIMIZING THE PREDICTABLE VARIANCE IN THE TWO SHIPBOARD CRITERION MEASURES

Although no measure of the amount of reliable variance was available for either shipboard criterion measure, it was considered of interest to determine the correlation between an optimum combination of the predictor scores and each of the criterion measures.

Variables for computing multiple correlation coefficients were selected using the Gengerelli approximation method. With this method, the values of the various regression coefficients are chosen successively until the pool of variables making independent contributions of valid variance is exhausted (3).

Job Performance Test

Table XIV shows the validities, beta weights, and multiple correlation coefficients for the best combination of predictors for the shipboard job

UNCLASSIFIED

performance test score. A combination of two work sample performance scores developed from the training environment and two basic aptitude test scores produced a multiple R of .<u>56</u>. Because cross-validation was not possible a standard correction for shrinkage was applied which gave a corrected R of .<u>53</u>.

Table XIV

Best Combination of Predictor Variables for the Separate Shipboard Criterion Measures

Criterion Measures	Variable	Validity	Beta	Multiple R
Shipboard Job Performance Test	Milling Operations, Basic	.37	.37)	
	Lathe Threading, Fine	.34	.24)	R = .56
	Perceptual Speed	.32	.29)	cR = .53*
	Spatial Relations	.32	.18)	
Shipboard Supervisory Ratings	Lathe Turning to Specs.	.40	.40)	
	MR School Achievement (written exams)	.35	.28)	R = .51 cR = .49*
	Lathe Threading, Fine	.32	.13)	

* Because cross-validation was not possible, a standard formula for shrinkage was applied: $cR = [1 - (1 - R^2)(\frac{N - 1}{N - m})]^{1/2}$ where N is the number of cases and m the number of variables.

<u>Supervisory Ratings</u>

A similar procedure was followed in selecting variables that would maximize the predictable variance in the shipboard rating scale scores. In this case two of the work sample scores from the training environment, plus the scores defined by the formal achievement tests at the MR School, provided the best combination of predictors. The computed multiple R was .<u>51</u> which corrected for shrinkage to .<u>49</u>. The validities and beta weights for each predictor variable may be inspected in Table XIV.

DEVELOPMENT OF A COMPOSITE SHIPBOARD CRITERION MEASURE

Examination of the relationships between scores on various traits of the shipboard rating scale and various aspects of performance on the shipboard performance

UNCLASSIFIED

test had revealed no systematic patterns of non-zero correlations. Furthermore, total score on the rating scale correlated very nearly zero with total score on the shipboard performance test.*

Since both shipboard criteria were predictable to a substantial degree, and in a logical way, by the various measures made during training, it was readily apparent that a combination of the two shipboard measures was justifiable and probably would represent a closer approximation of the ultimate criterion than either measure alone.

Since there appeared to be no logical reason for weighting either shipboard criterion measure more heavily than the other, a simple summation of the total rating scale score and the total job performance test score, after each had been reduced to centile form, was made. The resulting score is hereafter referred to as the composite shipboard criterion score.

CORRELATIONS OF THE COMPOSITE SHIPBOARD CRITERION WITH VARIOUS PREDICTORS

In Table XV, the correlations of the 20 predictor variables with the composite shipboard criterion are shown. It will be noted that each broad category of predictors, aptitude, work samples, formal school achievement, interest scores, and instructor ratings achieved one or more significant correlations with the composite criterion. In all, 10 of the 20 predictor variables correlated significantly with the composite with two others falling just short of the 5% significance level.

* It is of interest that, contrary to the findings of earlier studies, neither shipboard measure was related to the pay grade of the subjects. Pay grade correlated .04 with total score on the shipboard performance test and -.01 with total score on the shipboard rating scale. (N = 30 2nd Class MR's and 23 3rd Class MR's.)

UNCLASSIFIED

UNCLASSIFIED

Table XV

Correlations of Predictor Variables with the Composite Shipboard Criterion Measure

<u>Aptitude Factors</u>	
Verbal Comprehension	.16
General Reasoning	-.09
Number Facility	-.12
Mechanical Knowledge	.43**
Spatial Relations	.33**
Associative Memory	.01
Perceptual Speed	.30*
Motor Control	.00
<u>Interest Factors</u>	
Interest in Mechanical Rates	.27*
Interest in Mechanical Activity	.09
<u>Work Sample Factors</u>	
Lathe Threading, Basic	.26*
Lathe Threading, Fine	.44**
Hand Tool Performance	.23
Lathe Precision Turning	.22
Lathe Turning to Specs.	.31*
Lathe Turning to Finish	.12
Milling Operations, Basic	.37**
Milling Operations, Complex	.11
<u>Academic Achievement Factor</u>	
MR School Written Exams	.35**
<u>School Rating Factor</u>	
Rated Suitability as an MR	.42**

* significant at 5% level of confidence
** significant at 1% level of confidence

In addition to the fact that so many of the predictors correlated significantly with the composite shipboard criterion, it is of interest that <u>Interest in Mechanical Rates</u>, which had failed to correlate appreciably with either of the shipboard criterion measures taken separately, was related to the composite criterion to a significant degree.

The next procedure was to maximize the predictable variance in the composite shipboard criterion measure. For this, a more exact measure of the beta weights than those provided by the Gengerelli approximation was desired. The Doolittle method of solving simultaneous equations therefore was selected.

The variables to be included in the multiple correlation should be those having the highest validities and the lowest intercorrelation with other valid predictors. The matrix of intercorrelations of the valid predictors is produced in Table XVI for easy reference.

Table XVI

Intercorrelations of Predictor Variables
and Their Validities for the Composite Shipboard Criterion

	Mech. Knowl.	Spat. Rel.	Perc. Speed	Interest in Mech. Rates	Acad. Achiev. Factor	School Rating Factor	Lathe Threading Fine	Mill. Mach. Basic	Lathe Turning Specs.	Lathe Threading Basic
<u>Aptitude Factors</u>										
Mechanical Knowledge										
Spatial Relations	.47									
Perceptual Speed	.38	.09								
<u>Interest in Mech. Rates</u>	.39	.14	.04							
<u>Academic Achievement Factor</u>	.43	.40	.14	.17						
<u>School Rating Factor</u>	.38	.30	.19	.16	.63					
<u>Work Sample Factors</u>										
Lathe Threading, Fine	.24	.19	.21	.10	.18	.30				
Milling Machine, Basic	.24	.25	-.13	.17	.19	.23	.28			
Lathe Turning, Specs.	.22	.05	.04	.23	.17	.32	.36	.25		
Lathe Threading, Basic	.21	.13	.21	.22	.37	.45	.54	.20	.15	
Correlation of Each Predictor with Composite Shipboard Criterion Score	.43**	.33**	.30*	.27*	.35**	.42**	.44**	.37**	.31*	.26*

** significant at 1% level
* significant at 5% level

MAXIMIZING THE PREDICTABLE VARIANCE IN THE COMPOSITE SHIPBOARD CRITERION

In Table XVII, the validities and beta weights for the best combination of predictors for the composite criterion are shown. It will be noted that three work sample measures from the school environment, one aptitude measure, and the ratings of suitability given by the school instructors are involved. Using this combination of predictors, a multiple R of .<u>62</u> (corrected for shrinkage to .<u>58</u>) was obtained.

Table XVII

Best Combination of Predictor Variables for the Composite Shipboard Criterion Measure

	Variable	Validity	Beta	Multiple R
<u>Composite Shipboard Criterion</u> (Unweighted sum of total job performance test score and total supervisory rating score)	Lathe Threading, Fine	.44	.25)	
	Mechanical Knowledge	.43	.24)	
	Rated Suitability for the MR Job	.42	.19)	R = .62 cR = .58
	Milling Operations, Basic	.37	.18)	
	Lathe Turning to Specifications	.31	.06)	

The multiple correlation of .62 indicates that nearly 40% of the total variance in the composite shipboard criterion score could be accounted for the variance in scores obtained two years earlier on work samples produced in the training environment, a test of mechanical knowledge, and the predictions of school instructors of subsequent suitability as a machinery repairman.

Absence of reliability data for the criterion scores does not permit an estimate of how much of the reliable variance in the composite shipboard criterion was predictable by these variables. In view of known uncontrolled conditions that prevailed during the administration of the performance test aboard ship, and in view of the estimated split-half reliability for the performance test of .<u>87</u>, it is regarded as unlikely that the composite shipboard criterion had a reliability in excess of .75. If this value can be taken for the sake of argument, the predictors listed in Table XVII account for over 50% of the true variance in the shipboard criterion. Even allowing for the shrinkage that would be expected with cross-validation, the extent to which the variance in the composite shipboard criterion has been accounted for provides considerable insight into the nature of a suitable criterion for the shipboard performance of mechanical rates in the Navy.

UNCLASSIFIED

Chapter 5

DISCUSSION

The findings of this study suggest that the shipboard performance of certain Navy rates can be meaningfully assessed by means either of job sample performance tests or supervisory ratings. Most important, however, they indicate that both kinds of measures may be required if the most meaningful shipboard criterion is to be developed.

These conclusions are reached largely on the basis of the essential independence observed in the present study between ratings assigned by supervisors aboard ship and performance aboard ship as measured by a job sample test. Many other studies, both by the present investigators and others, similarly have produced low correlations between ratings and performance tests. While such results often have led to questions about the utility of either type of measure as a practical criterion, abundant evidence was produced by the present study to indicate that both measures may bear substantial relationships to a number of logically pertinent predictor variables and still remain largely independent of each other. The use of both, therefore, may not only be desirable, but essential if a proper appraisal of performance is to be obtained.

Other arguments can be advanced from past research concerning the probable complementary nature of these two criterion measures. The need to employ practical performance tests seems certain in view of the almost universal surprise of officers, petty officers, and instructors when they observe the actual performance capability of their men under controlled test conditions. Again and again the performance capabilities of Navy personnel, as tested, have fallen below the expectations of observers who felt they knew the skills of their men very well. This has been found true in both the training and operational environment.

The need for supervisory ratings as a criterion measure to complement tested performance also can be supported on the basis of observations from past studies. For one thing, it is known that the factorial structure of ratings is complex. For another, it is evident that a number of important intangibles, which may include adaptation to shipboard life, good relations with superiors and other men, performance of military duties, ability to understand and communicate instructions, willingness and ability to learn, and so forth, very probably are reflected to some degree in the ratings assigned by superiors.

The ratings assigned by MR School instructors two years prior to this follow-up study clearly contained variance of an intellectual and academic sort as well as those kinds associated with the technical skills required to produce fine, machined components. For this reason, it would seem that the predictions of Class "A" School instructors, gathered with a properly designed rating scale and the proper training in rating methods, might serve as an important intermediate criterion for many studies concerned with the selection, classification, and placement of enlisted Naval personnel. Generalizing from one mechanical rate to others may not be fully justified, of course. It seems reasonable, however, that such a criterion would be superior to the frequently used school grades with their high proportion of verbal variance. It can be said without fear of contradiction that when written test scores are used as criteria, verbal and other kinds of intellectual variance will play a much more important role than they apparently do in subsequent performance aboard ship.

Both practical performance tests and ratings of performance by superiors should be the subject of continued investigation by those studying performance criteria. The nature of performance tests can be largely specified and determined by those skilled in performance test construction. The nature of performance ratings, and the bases upon which one human being judges the performance and the

UNCLASSIFIED

desirability of others in the job environment, remain unclear in spite of the tremendous number of studies where performance ratings have been employed. This is likely to remain the case until judgments of human performance are made under more controlled conditions and under circumstances where they can be related to objective external criteria. It is to this problem that future investigations under this contract will be directed.

BIBLIOGRAPHY

1. Buckner, D. N. Research on the development of performance criteria, Technical Report VIII. The predictability of ratings as a function of inter-rater agreement. Los Angeles, Human Factors Research, Incorporated, 1957. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).

2. Catell, Raymond B. Factor analysis. New York: Harper and Brothers, 1952.

3. Gengerelli, J. A. "A simplified method for approximate multiple regression coefficients", Psychometrika, 1943, 3, 135-146.

4. Guilford, J. P. Fundamental statistics in psychology and education, New York: McGraw-Hill, 1950.

5. Guilford, J. P. (ed.) Printed classification tests, Report No. 5, Army Air Forces Psychology Program Research Reports, 1947.

6. High, W. S. and Mackie, R. R. Research on the development of performance criteria, Technical Report VII: A factor analytic study of aptitudes, interests and practical performance skills of Navy machinery repairman students. Los Angeles, Human Factors Research, Incorporated, 1957. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).

7. Mackie, R. R. Several treatments of inter-group differences and their effect on the reliability and predictability of performance ratings, Technical Report No. 3. Submarine personnel selection. Los Angeles, Human Factors Research, Incorporated, 1955. (For Bureau of Medicine and Surgery, United States Navy and the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1113(00).

8. Mackie, R. R., Strayer, F. K. and Buckner, D. N. Validation of the U. S. Medical Research Laboratory personnel assessments, Technical Report No. 1. Submarine personnel selection. Los Angeles, Human Factors Research, Incorporated, 1954.* (For Bureau of Medicine and Surgery, United States Navy and the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1113(00).

9. Mackie, R. R., Wilson, C. L. and Buckner, D. N. Research on the development of shipboard performance measures, Part V: Interrelationships between aptitude test scores, performance in submarine school, and subsequent performance in submarines as determined by ratings and tests. Los Angeles, Human Factors Research, Incorporated, 1954. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00)

10. Mackie, R. R., Wilson, C. L., Buckner, D. N., De Loach, W. D., Dossett, W. F. and Ross, E. B. Practical performance test batteries for electrician's mates and radiomen developed in conjunction with A manual for use in the preparation and administration of practical performance tests. Los Angeles, Human Factors Research, Incorporated, 1953. (For Bureau of Naval Personnel (Pers 15), United States Navy and Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 876(00).

* Reports published prior to 1955 were issued under the firm's former name, Management and Marketing Research Corporation.

UNCLASSIFIED

UNCLASSIFIED

11. Wilson, C. L. and Mackie, R. R. Research on the development of shipboard performance measures. Part I: The use of practical performance tests in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, Human Factors Research, Incorporated, 1952. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy.) Contract Nonr 1241(00).

12. Wilson, C. L., Mackie, R. R. and Buckner, D. N. A manual for use in the preparation and administration of practical performance tests. Los Angeles, Human Factors Research, Incorporated, 1953. (For Personnel Analysis Division, Bureau of Naval Personnel and the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Washington 25, D. C.) Contract Nonr 876(00). This was done in collaboration with Siegel, A. I. and Courtney, D. of the Institute for Research in Human Relations.

13. Wilson, C. L., Mackie, R. R. and Buckner, D. N. Research on the development of shipboard performance measures. Part II: The use of a performance rating scale in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, Human Factors Research, Incorporated, 1954. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).

14. Wilson, C. L., Mackie, R. R. and Buckner, D. N. Research on the development of shipboard performance measures. Part III: The use of performance check lists in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, Human Factors Research, Incorporated, 1954. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).

15. Wilson, C. L., Mackie, R. R. and Buckner, D. N. Research on the development of shipboard performance measures. Part IV: A comparison between rated and tested ability to do certain job tasks. Los Angeles, Human Factors Research, Incorporated, 1954. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).

11. Wilson, C. L. and Mackie, R. R. Research on the development of shipboard performance measures. Part I: The use of practical performance tests in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, Human Factors Research, Incorporated, 1952. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy.) Contract Nonr 1241(00).

12. Wilson, C. L., Mackie, R. R. and Buckner, D. N. A manual for use in the preparation and administration of practical performance tests. Los Angeles, Human Factors Research, Incorporated, 1953. (For Personnel Analysis Division, Bureau of Naval Personnel and the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Washington 25, D. C.) Contract Nonr 871(00). This was done in collaboration with Siegel, A. I. and Courtney, D. of the Institute for Research in Human Relations.

13. Wilson, C. L., Mackie, R. R. and Buckner, D. N. Research on the development of shipboard performance measures. Part II: The use of a performance rating scale in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, Human Factors Research, Incorporated, 1954. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).

14. Wilson, C. L., Mackie, R. R. and Buckner, D. N. Research on the development of shipboard performance measures. Part III: The use of performance check lists in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, Human Factors Research, Incorporated, 1954. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).

15. Wilson, C. L., Mackie, R. R. and Buckner, D. N. Research on the development of shipboard performance measures. Part IV: A comparison between rated and tested ability to do certain job tasks. Los Angeles, Human Factors Research, Incorporated, 1954. (For Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Department of the Navy) Contract Nonr 1241(00).