RESEARCH ON THE DEVELOPMENT OF SHIPBOARD PERFORMANCE MEASURES

FINAL REPORT
(IN FIVE PARTS)

PART II

THE USE OF A PERFORMANCE RATING SCALE IN THE MEASUREMENT OF SHIPBOARD PERFORMANCE OF ENLISTED NAVAL PERSONNEL

PREPARED FOR

PERSONNEL AND TRAINING BRANCH
OFFICE OF NAVAL RESEARCH
DEPARTMENT OF THE NAVY

BY

MANAGEMENT AND MARKETING RESEARCH CORPORATION
LOS ANGELES, CALIFORNIA

RESEARCH ON THE DEVELOPMENT OF SHIPBOARD PERFORMANCE MEASURES

FINAL REPORT
(In Five Parts)

PART II

THE USE OF A PERFORMANCE RATING SCALE IN THE MEASUREMENT OF SHIPBOARD PERFORMANCE OF ENLISTED NAVAL PERSONNEL

By

Clark L. Wilson, Ph.D.

Robert R. Mackie, Ph.D.

Donald N. Buckner

February 1954

Prepared for
Personnel and Training Branch
Office of Naval Research
Department of the Navy
Washington 25, D. C.

Contractor
Management and Marketing
Research Corporation
3004 South Grand Avenue
Los Angeles 7, California
for
Psychological Research Center
Los Angeles, California

Contract N8 onr 70001

ACKNOWLEDGEMENTS

We wish to express our sincere thanks for the cooperation of many Naval personnel, afloat and ashore. Particularly, we wish to thank Capt. G. R. Donaho, USN, for permission to begin the work afloat in Submarine Flotilla One, U. S. Pacific Fleet, Subsequent Commanders of Submarine Flotilla One have been most helpful in the continuing research.

Force, Squadron, Division, and Tender staffs in both Pacific and Atlantic Fleets have contributed greatly to the progress of the work, and our especial appreciation goes to the Officers and leading Petty Officers aboard the more than thirty submarines where the work was done.

Capt. T. L. Willmon, MC, USN, and Commander Gerald J. Duffner, MC, USN, as Officers-in-Charge, Medical Research Laboratory, U. S. Naval Submarine Base, New London, have both done much to further the work of this project.

We also wish to acknowledge the editorial assistance of Forrest K. Strayer and the contributions of William F. Dossett and Wells D. De Loach during the data analysis phase of the project.

TABLE OF CONTENTS

ABSTRACT

A performance rating scale that included 10 traits reflecting various technical and non-technical aspects of Navy shipboard performance was developed and used to evaluate the performance of Electrician's Mates (EM's) and Enginemen (EN's) serving aboard submarines in the Atlantic and Pacific Fleets. Analysis of results indicated that officers and petty officers using the scale tended:

(1) to agree with one another when they evaluated the same men

(2) to be consistent in their own evaluations from one time to the next

(3) to discriminate reliably among men of the same pay grade

(4) to differentiate, to an appreciable degree, the technical from the adjustive aspects of shipboard performance.

In addition, a factorial analysis indicated that:

(5) at least two broad "factors" of shipboard performance - one representing technical skill, and the other, adjustment to Navy life - accounted for most of the intercorrelations among traits

(6) the traits representing the technical side of performance correlated moderately high with independent measures of technical skill, but the traits representing the adjustment side of performance were not related appreciably to any other measures obtained.

As a part of this overall research project practical performance tests and performance check lists for EM's and EN's were also developed (see Parts I and III of this Final Report for details).

Relationships between these two measures and the rating scale are reported and discussed in this report.

Chapter I

SUMMARY AND OPERATIONAL IMPLICATIONS

At the time this research was initiated, shipboard performance in the Navy was evaluated by means of Quarterly Marks assigned by division officers. It was considered by many, in the Navy and out, that the Quarterly Marks did not adequately reflect the differences among enlisted men's abilities to perform their duties. There was a marked tendency for the assigned scores to pile up around the highest possible value (4.0). Unless it were assumed that most men's performance was of equal excellence, these marks were failing to give discriminating and realistic indications of performance.

For this reason, research was conducted to determine if more adequate means of rating shipboard performance could be developed. This report, which is Part II of a Final Report of a study of shipboard performance measures, describes a performance rating scale that was developed for use by submarine officers and petty officers in evaluating the performance of Electrician's Mates (EM's) and Enginemen (EN's) serving under them.

In its final form, the performance rating scale, following a format that permitted man-to-man comparisons, included these ten traits:

Cooperation	Neatness of Work
Knowledge of the Job	Care of Equipment
Application and Initiative	Ability to Troubleshoot
Judgment and Common Sense	Sincerity in Doing a Good Job
Leadership	Discipline

In addition, a Rank Order Preference (relative overall value in the gang) was included as a summary evaluation.

A total of 320 EM's and 487 EN's were rated on these traits. The majority

of the men were rated by at least three officers and/or petty officers. The reliability (consistency) of the ratings was calculated by having a representative group of raters re-rate their men after a suitable lapse of time. It was found to be quite high (r_t = .88). Raters also agreed with one another substantially when they rated the same men, the average inter-rater agreement on total score being r_t = .70.

The ratings assigned by all raters on each of the traits were intercorrelated and analyzed to find what general elements (factors) of performance could account for the intercorrelations. Two factors were isolated in each of two independent analyses:

1. Technical competence (ability to meet the practical or technical demands of a job)
2. Personal adjustment (attitude towards both job and other elements of Navy life).

The measures obtained with the performance rating scale were also correlated with scores from two other measures of shipboard performance: performance check lists and practical performance tests. The check lists were designed to provide ratings and the performance tests objective scores of ability to perform specific tasks from EM's and EN's jobs.

It was found that the rating scale and check list scores were highly correlated (about .85) when men in all pay grades were rated together (the substantial correlation between pay grade and performance ratings in part induced this correlation), but the correlations between rating scale trait scores and performance test scores were low to moderate (.25 to .62). On a within-pay-grade basis, the two rating devices correlated moderately (.23 to .61) while the rating scale scores and performance test scores correlated low to moderately (.02 to .40). The

correlations were generally higher for the EM's than for the EN's.

It was concluded that discriminating and reliable ratings of general traits can be obtained which will reflect important aspects of shipboard performance. Such ratings may not, however, indicate very much about ability to perform specific tasks from shipboard jobs.

The present and probable continued use of rating devices as measures of shipboard performance gives operational significance to many of the findings of this study.

Specifically:

1. Reliable and discriminating ratings can be made by Navy officer and petty officer personnel.
2. A man-to-man rating format appears to promote discriminating ratings and inter-rater agreement.
3. Ratings should be performed separately for different gangs and pay grades.
4. Both technical and adjustment aspects of shipboard performance can be rated.
5. Inter-rater differences in leniency and spread of ratings must be accounted for in order to achieve maximum usefulness from ratings.
6. Comparisons of the technical skill of men from different boats demand the use of performance tests as well as ratings.

Chapter II

BACKGROUND AND DEVELOPMENT OF THE RATING SCALE

The problem

"In wartime ... elaborate routines of test administration aboard ship may not be feasible, and a more convenient procedure for evaluating performance is needed. The use of rating methods would provide the means for making such evaluations."[1]

The problem of evaluating the shipboard performance of Navy men is an exceedingly important one. The success and constant improvement of selection, training and placement programs depend in large part on the availability of suitable performance measures which can be used as criteria against which the effectiveness of these programs can be judged.

In the Navy, and particularly in the submarine Navy, evaluations of performance require consideration of a complexity of factors. Technical know-how and skill are essential; even more important, however, may be factors of an attitudinal or adjustive nature toward Navy life and the men with whom one lives 24 hours a day.

This study is concerned with an effort to devise a rating scale which could be used to obtain measures of the many important aspects of shipboard performance. It, together with other studies on the problem of shipboard performance measures, provides partial answers to this important problem.[2]

---

[1] Human Factors in Military Efficiency, Summary Technical Report of the Applied Psychology Panel, NDRC, Vol. 1, Washington, D. C., 1946, p. 180.

[2] See also Part I, "The use of practical performance tests in the measurement of shipboard performance of Naval enlisted personnel," and Part III, "The use of performance check lists in the measurement of shipboard performance of enlisted Naval personnel," of this final report.

The Quarterly Marks System

At the time this research was initiated, shipboard performance was regularly evaluated in the Navy through the assigning of Quarterly Marks. Enlisted men were rated on their Proficiency in Rate and Conduct, and, depending on their job and pay grade, on their Leadership, Seamanship, and Mechanical Ability.

The Quarterly Marks scale had a score range from 0.0 to 4.0. However, a mark below 2.5 indicated unsatisfactory performance, and there was a tendency, in practice, for both Division Officers and the men being rated to consider any mark below 3.5 unsatisfactory. As a result, the marks tended to pile up between 3.5 and 4.0, with the majority for the higher pay grades at or near 4.0. Little discrimination resulted from such ratings and a man's scores seldom gave a clear indication of whether the job he was doing was excellent, good, ordinary, or just fair.

The problems faced by the Division Officer (or whoever does the rating) in assigning Quarterly Marks may be illustrated by a few extracts from the BuPers Manual:

> In determining marks, it shall be borne in mind that the mark for proficiency in rate is intended to be sufficient in itself to denote a person's ability, habits, and character; in short, the individual's value to the service in the particular rate.[3]

The following qualifications are then listed as the determining factors in assigning marks of 4.0 and 3.5:

4.0 Competent and reliable in rate. Not less than 3.5 in conduct.

3.5 Competent and qualified in all duties of rate; has qualities sufficient to justify advancement. Not less than 3.5 in conduct.[4]

---

[3] Bureau of Naval Personnel Manual, pp. 181-182.

[4] Ibid., p. 183.

This evident lack of distinction between the marks, and the emphasis put on a single, broad, performance characteristic may account, in part, for the fact that most of the Quarterly Marks were at, or near, 4.0.

With this brief critique of the characteristics of Quarterly Marks, let us pass on to the considerations which governed the design of the rating scales developed for this study.

Considerations for the design of a rating scale

A rating scale is a device that should help an officer or petty officer organize and express his evaluations of the men who work under him. In order that they be useful, such evaluations depend to a great extent on three things:

1. The format of the scale. Does it provide reference points by which the rater can reliably determine where each man should be placed on the scale?
2. The way relevant aspects of performance are described. Do they help the rater "break down" the job? Do they bring out the different features of performance? Do the trait descriptions help the rater visualize the man on the job?
3. The way "levels of performance" are defined. Do they help the rater separate the "good" from the "not so good", and the "superior" from the "above average"?

Requirements of rating scale construction and design

Investigation of the rating scale literature and consideration of the goals of this research have led to the following general requirements which affect rating scale construction and design:

1. A rating scale should be designed to reflect, insofar as possible, the

differences among the performance levels of the men being rated. That is, ratings given to a group of men should be spread over a relatively large portion of the scale and few men, if any, should receive the same score. The scale should yield scores that discriminate.[5]

2. A rating scale should be designed so that the "average rating" of a typical group of men falls somewhere near the middle of the scale. The bunching of scores at the upper end of the scale, so typical of ratings, is unrealistic and reduces discrimination as well. (It will be remembered that the Quarterly Marks had a tendency to be bunched at the upper end of the scale. This was a fault of the scale design as well as a result of overly lenient rating practices.)

3. A rating scale should be designed so that different raters will agree on the position occupied on the scale by each man rated. This is very difficult to achieve when raters are forced to rely on descriptive adjectives or numerical values in establishing key points on the scale. More complete verbal descriptions of observable bits of behavior should be of some help to raters in agreeing on the positions their men occupy. However, the format of the rating device can also be such as to promote agreement.

4. The design of a rating scale should be such that it will help the raters using it to be consistent in their ratings of the same men from one rating period to the next. This is reliability in the test-retest sense. Assuming raters had no recollection of scores given at an

---

[5] This assumes, of course, that men are actually different from each other in the performance characteristics on which they are to be rated. From present day knowledge of individual differences this appears a reasonable assumption.

earlier time, there still should be a high relationship between scores given to a group of men at intervals of, say, 90 to 180 days.

With these requirements in mind, the actual construction of a performance rating scale was begun.

Format of the experimental scale

Rating scales typically are single-page arrangements with the name of the man to be rated at the top of the page and a series of traits on which he is to be rated listed down the side. Performance levels on this type of rating form are usually expressed in terms of some scale of numbers or adjectives, or both. For example, a man may be rated as "excellent" or "4.0" on one characteristic, or "good" or "3.5" on another, and so forth. This kind of rating form, though simple in design and easy to use, minimizes the likelihood of obtaining useful ratings. It fails to provide the raters with a satisfactory framework for comparison. The adjectives or numbers used usually have quite different meanings to different raters at different times. Either, or both, of these conditions leads to unreliable ratings. Words and numbers on a rating scale lack absolute meaning and are thus subject to as many interpretations as there are raters.

It is exceedingly difficult to state how absolutely good or poor a man is in his work. It is much easier to say he is good compared to this or poor compared to that. The more concrete and meaningful are the reference points against which these necessary comparisons must be made, the more objective and reliable will be the resulting ratings. Something more than numbers or adjectives is needed although they may be of limited help. The history of ratings suggests that this something more may be achieved by evaluating a man's performance relative to that of his peers.

In most job situations, men belong to an identifiable work group. Each man

within the group either does work similar to that of the others, or works under approximately the same conditions, or both. The members of this group provide meaningful, stable points of reference for all raters who are sufficiently well acquainted with the group to do the rating.

A man's assigned rank in a group on any one trait, or his average rank in a group on all traits, is an indication of his level of performance based on meaningful reference points -- namely, the other members of the work group.

The performance rating scale was designed with this man-to-man comparative procedure in mind.[6] The following features were considered important: (A copy of the final form of the rating scale may be seen in Appendix B.)

1. Space was provided on each page of the rating scale booklet for the names of all men within a given work group. Each rater rated each man as a member of a particular work group. In this study the group was either the EM or EN gang aboard submarines.[7]
2. A single performance trait was listed on each page of the booklet. All men were therefore rated as a group on ONE trait at a time.

   This procedure of rating a group of men on each trait separately is to be contrasted with the usual procedure of rating one man on all traits without direct reference to the other persons to be rated.
3. The definition of each trait was followed by four series of statements down the side of the page defining various degrees of possession of the

---

[6] The design was not completely original. It was first tested by Stevens and Wonderlic, "An Effective Revision of the Rating Scale Technique," Personnel Journal, 13; 1934, 125-134. It was recommended by Guilford, Psychometric Methods McGraw-Hill Book Co., New York, 1936, 263-264. It was further tested by Gilinski "The Influence of the Procedure of Judging on the Halo Effect," American Psychologist, 2:309-319, 1947.

[7] In subsequent work, groups have been homogeneous with respect to pay grade as well as rate.

trait in question. The top-most group of statements described exceptionally outstanding behavior. The next group described clearly above average performance; the third group described average or satisfactory performance and the last group of statements described below average or undesirable performance.

Thus, three of the four groups of statements described acceptable performance or better. This was done in an effort to keep the average rating down toward the physical center of the scale, recognizing both the tendency of raters to rate their men high and, at the same time, the natural selection and admittedly superior performance of most higher rated petty officers.

4. In a column underneath each man's name was a continuum line. This line was unbroken except for short horizontal marks which located on the continuum the position of each group of descriptive statements. The raters could thus locate where, on the line, a given description of behavior would fall. They could then decide whether the performance of a man being rated was equal to, better or worse than a particular description and rate him at any appropriate point along the entire continuum.

On each page of the rating scale booklet, therefore, appeared a definition of the trait on which the men were to be rated, four statements defining the degree of possession of the trait, and the names of the men being rated. Below each of the latter was a column, representing a continuum, in which the ratings were to be made with a check mark.

Each time a man was rated on a given trait, the ratings assigned other men had to be taken into consideration by the

rater. Thus, each man was directly compared with each other man in the group on each of the traits.

The selection of performance traits for the rating scale

To explore as many facets of shipboard performance as possible, a wide selection of traits was made for the preliminary form of the rating scale. These traits, which are listed below, were gleaned from a study of shipboard jobs and from conversations with Naval personnel:

Social Adjustment	Adaptability
Quality of Work	Leadership
Neatness of Appearance	Overall Efficiency
Cooperation	Neatness of Work
Watch Standing	Ability to be Taught
Knowledge of the Job	Care of Equipment
Discipline	Ability to Troubleshoot
Application and Initiative	Sincerity in the Job
Dependability	Manual Skill

Overall Efficiency in Rate (within pay grade)

In addition to an overall rank order preference, the final form of the rating scale consisted of ten traits. These were:

Cooperation	Neatness of Work
Knowledge of the Job	Care of Equipment
Application and Initiative	Ability to Troubleshoot
Judgment and Common Sense	Sincerity in Doing a Good Job
Leadership	Discipline

Details of the analysis that led to this final selection of traits are reported in Chapter III.

Use of the performance rating scale aboard ship

As an exploratory study, the preliminary rating scale was used to obtain evaluations of 187 EM's and EN's of Submarine Squadrons 3 and 7 at San Diego.

All pay grades from strikers through chief petty officers (CPO's) were included. The men rated were selected solely on the basis of availability.

A revised preliminary form was used later to obtain evaluations of 286 submarine EM's and EN's. Analysis of these two preliminary forms led to the development of the final form which is included in Appendix B to this report.

The final form of the rating scale was used to evaluate the same groups of EM's and EN's that were subsequently evaluated with the performance check lists and tested with the practical performance tests described in Parts I and III of this report. These groups included 320 EM's and 487 EN's, the greater majority of the men being strikers and third class petty officers.

An effort was made to get as many raters as possible who felt they knew the men to be rated well enough to rate them fairly. In most instances, this meant the Engineering Officer, the two leading Chief EM's, and the two leading Chief EN's. In some instances, ratings were made by the Executive Officer or Assistant Engineering Officer.

The majority of the EM's and EN's were rated by either one officer and two petty officers or by two officers and one petty officer; so, typically, at least three ratings were obtained on each man.

Instructions to the raters

Written instructions, together with sample ratings, were given to all raters. (Examples may be seen on pages 1 and 2 of the rating scale in the Appendix.) In addition, in the preliminary study raters were instructed personally by project personnel. Common pitfalls of rating, such as excessive leniency, halo effect, and high interrelationships among traits due to logical errors were explained to the raters in simple terms and they were asked to

guard against them. During later administrations, time considerations prevented personal instruction and the written instructions had to suffice.

The written instructions emphasized the following:

1. The differences, no matter how small, in the performance levels of the men being rated should be brought out. Men in each work group should be compared with one another on each trait. It was suggested that the poorest and best man in each group ("best" or "poorest" on the trait in question) should be rated first, and then the other men rated in between. Raters were instructed to avoid giving tie ratings if at all possible.
2. The men's pay grades should not be unduely influential in the in the ratings assigned. Men were to be rated on each characteristic strictly on the basis of their demonstrated performance. Raters were cautioned against the general "halo-effect" of pay grade.

The halo created by pay grade is a real problem in analyzing ratings given to Naval personnel. It is natural for an officer or petty officer to be affected by the man's pay grade when he evaluates a man's performance. Part of this influence, no doubt, is the result of genuine differences in ability reflected in the different pay grades, but part very likely is spurious.

The effect of pay grade on the usefulness and reliability of ratings was examined in considerable detail during this study. Most statistical analysis reported in later chapters were performed both on a within and across pay grade basis in order to answer the many questions involved.

Scoring the ratings

It will be remembered that each definition of a performance trait on the pages of the rating scale booklet was accompanied by four groups of statements down the side of the page defining the degree of possession of that trait. Across the top of the page were listed the names of the men to be rated. A check mark in a column opposite one of the series of qualifying statements or between two such series, and under a particular man's name, indicated that man's rated performance level on a particular trait.

The distance of the check mark from the lower end of the column was converted into a numerical score by measuring it with a centimeter rule. All columns were thirteen centimeters long, and so possible scores ranged from 0.0 to 13.0. Each check mark was read to the nearest half centimeter.

These raw scores varied a good deal from rater to rater in central tendency and spread. To make ratings from different raters comparable, it was necessary to convert them into standard scores.[8] This had the effect of producing a common average for all ratings and reduced the effect of differential spread.

[8]Standard scores were expressed in terms of the Sten-scale. Scores on this scale ranged from 0 to 9 inclusive, with the mean at 4.5. Each unit on the scale covered one-half standard deviation.

Chapter III

RESULTS FROM SHIPBOARD USE OF THE RATING SCALE

A. Measuring Characteristics of the Scale

Average rating and spread of ratings over the scale

In Table I, appearing on page 17, the mean and standard deviation of ratings are reported for each trait of the preliminary form of the rating scale. This analysis was not repeated for the revised form, there being no suggestion of any differences in results. If anything, the variability of ratings on the revised form probably was slightly greater than that reported here because of the specific instruction in later administrations to avoid tie ratings. To indicate the variability of the means and standard deviations, the middle 80% range of values has been included in the Table.

In many rating scales there is a notable tendency for raters to reduce the usefulness of their evaluations by:

(1) Assigning high ratings, on the average

(2) Indicating little or no differences between men (small dispersions).

As indicated in the previous chapter, it was thought that the format of this scale, together with the manner in which the trait descriptions were graduated, would reduce both these tendencies appreciably.

In large part these expectations were realized. There was, of course, variability in the leniency - stringency factor, mean ratings for individual raters extending all the way from 5.9 to 10.0 on the 13-point scale. However, the mean of mean ratings for 43 raters in the preliminary study was 7.8 -- quite acceptably near the center of the continuum, 6.5.

In general, the dispersions of ratings were sufficiently large to provide meaningful and useful discriminations. There were a few raters who assigned a large percentage of tie ratings, thus producing small standard deviations and little discrimination. This led to the instruction, for later administrations, to avoid tie ratings if at all possible.

Considerable variability can be seen in the means and spread of scores from trait to trait. This could be due to one of several factors or in part due to all of them. One possibility is that the variability resulted from differences in the extremeness of phraseology of the descriptive phrases for the several traits. A second possible explanation is that raters may have considered it more logical or essential to assign high mean ratings for some traits than others. For example, a rater might reason that it was perfectly satisfactory to assign moderate or even low ratings in a trait such as Manual Skill, but might consider it a reflection on his own performance if anything but high ratings were assigned in a trait such as Discipline. Finally, a third possible explanation of differences in means and variabilities of the several traits is that they reflect differences which, in fact, do exist. For example, the mean score in Leadership may be low because a substantial proportion of the men rated actually have had no opportunity to display leadership. The mean rating in Discipline may be high because the majority of men offer no disciplinary problems. The variability in Manual Skill may be small because there is little opportunity to demonstrate differences in motor skills in the particular rates studied.

TABLE I

MEAN AND MIDDLE 80 PERCENT RANGE OF RATINGS, MEAN AND MIDDLE 80 PERCENT RANGE OF STANDARD DEVIATIONS FOR EACH TRAIT (Figures represent results from 43 raters aboard 10 submarines and a total sample of 187 Enginemen and Electrician's Mates)

Trait	Mean of all raters' means	Middle 80 percent range of means	Mean of all raters' S.D.'S	Middle 80 percent range of S.D.'S
1. Social Adjustment	7.2	5.8 - 9.6	1.8	0.6 - 2.6
2. Quality of Work	7.9	5.6 - 10.0	1.9	0.9 - 2.9
3. Neatness of Appearance	7.3	4.7 - 8.8	1.9	0.2 - 2.9
4. Cooperation	8.1	6.4 - 9.5	2.1	1.0 - 3.2
5. Watch Standing	7.4	5.2 - 9.4	2.1	1.1 - 3.0
6. Knowledge of Job	7.4	5.8 - 9.4	2.3	1.2 - 3.5
7. Discipline	9.5	7.8 - 11.0	1.9	0.8 - 3.2
8. Application and Initiative	6.6	4.8 - 8.8	1.9	1.0 - 3.0
9. Judgment and Common Sense	8.0	6.2 - 10.2	1.9	0.9 - 2.9
10. Dependability	7.2	4.9 - 9.4	2.1	1.3 - 3.1
11. Adaptability	8.7	6.9 - 10.4	1.9	0.9 - 3.1
12. Leadership	6.1	4.1 - 7.9	2.3	1.2 - 3.4
13. Overall Efficiency	8.4	6.5 - 10.0	2.0	0.8 - 3.3
14. Neatness of Work	7.2	5.2 - 9.2	1.8	0.8 - 2.7
15. Ability to be Taught	8.4	6.6 - 10.2	1.9	0.8 - 3.2
16. Care of Equipment	8.6	6.5 - 10.5	1.8	0.8 - 3.1
17. Ability to Troubleshoot	7.3	5.6 - 9.2	2.0	1.2 - 3.3
18. Sincerity in Job	8.5	6.8 - 10.2	2.0	0.8 - 3.5
19. Manual Skill	7.6	6.1 - 9.4	1.7	0.3 - 2.7
20. Overall Efficiency in Rate	8.6	7.1 - 10.2	1.9	0.1 - 3.2
Overall Means	7.80	5.93- 9.66	1.96	.84- 3.09

Reliability of the ratings

An important feature of any evaluating instrument is the consistency with which it measures. Few reports on rating scales of the traditional type indicate satisfactory reliability and, in many studies, reliability in the rate-rerate sense is not even reported.

Having established that raters could assign meaningful and discriminating ratings, it was next desired to ascertain how consistently (reliably) they did this.

One hundred fifteen subjects were common to the two preliminary administrations of the rating scale and also had the same raters each time. These were non-selected cases from six submarines, being singled out solely because of participation in both trial runs. The elapsed period of time between the first and second ratings varied from five to as much as nine months for the various raters.

Since fairly high reliability would be expected on the basis of knowledge of a man's pay grade alone, it was decided that the reliability should be estimated on a <u>within pay grade</u> basis. That is, for each pay grade, it was determined whether a given rater placed his men in essentially the same order (based on total score) during the second evaluation as he did on the first. The self-agreements and disagreements then were plotted in a four-fold table (split at the median) and the correlation (tetrachoric) was computed. For the 115 cases (over 300 ratings) the correlation was $r_t = .88$. This compares very favorably with reliability figures usually obtained with more objective measuring devices, such as psychological tests, and is considerably higher than the reliabilities usually attributed to rating scales (.60 to .70).

Agreement among the raters

In addition to reliability in the rate-rerate sense, a useful rating device must promote substantial agreement among raters on the relative proficiency of the men they are rating. Indeed, ratings probably can be considered a satisfactory criterion of performance in direct proportion to the amount of agreement among raters, provided, of course, the characteristics of performance being rated are relevant to the ultimate criterion.

Inter-rater agreement was determined by correlating one rater's judgments on each trait with those of each other rater who had rated the same subjects. Thus, if a pair of raters both considered a man either above or below the median on a certain trait, it constituted an agreement. If, on the other hand, one rater considered the man above the median and another considered him below it, a disagreement was recorded. By considering the regression of each rater on each other rater, a four-fold table of frequencies representing agreements and disagreements was established for each trait in the scale and for average score. From these four-fold tables, tetrachoric correlation coefficients were computed which served as indices of agreement.

The question again arose as to the role of pay grade in determining the magnitude of inter-rater agreement. For example, it seemed a reasonable assumption that most raters would generally agree that a first-class petty officer was better than a third-class petty officer or a striker. Thus the factor of pay grade alone would create considerable agreement among raters. It was considered necessary, therefore, to determine whether there was any basis for agreement among raters independent of the factor of pay grade.

Each rater's subjects were grouped by pay grade. On each trait these pay-grade groups were divided into upper and lower sub-groups with the division at the median. (If there was only one subject in any given pay

grade, he was omitted from the calculation.) Those subjects in the upper sub-group were assigned a passing score (+) while those in the lower sub-group received a failing score (0).

With these scores assigned for each rater for each trait for each pay grade, the intercorrelations of raters' judgments was again a matter of plotting a four-fold table of frequencies. All pay grades were plotted on the same table and each subject appeared in the table as many times as there were pairs of raters who judged him.

The coefficient computed from this table may be termed a within pay grade coefficient. The results for each trait, under each of the two conditions, and for both of the preliminary samples, are shown in Table II on the following page. In addition to the agreement on the single traits, the agreement on average scores is indicated. The values for the traits of Neatness of Appearance, Dependability, Adaptability, Ability to Troubleshoot, and Manual Skill do not appear for the second sample since those traits were eliminated from the second form of the scale.

Several of the results reported in Table II require comment. There is considerable variability in the size of the coefficients from trait to trait as might be expected. Considering both samples and both conditions of correlation, the trait Social Adjustment seems to be about the least objective. This is reasonable in view of the obvious difficulties involved both in defining Social Adjustment and in agreeing on objective indicators for such a trait.

Among the traits leading to the highest degree of agreement under both conditions of correlation were Knowledge of the Job, Discipline, Judgment and Common Sense and Leadership. It is felt that the behavioral referents for

TABLE II

INTER-RATER AGREEMENT* ON THE VARIOUS RATING SCALE TRAITS AND ON AVERAGE SCORE, FOR BOTH SAMPLES, WITH PAY GRADE IN AS A VARIABLE AND WITH PAY GRADE HELD CONSTANT

Trait	First Sample (N=107)		Second Sample (N=206)	
	Pay Grade In	Pay Grade Out	Pay Grade In	Pay Grade Out
1. Social Adjustment	.32	.26	.46	.40
2. Quality of Work	.53	.56	.55	.39
3. Neatness of Appearance	.42	.42	-	-
4. Cooperation	.30	.58	.50	.57
5. Watch Standing	.45	.44	.60	.59
6. Knowledge of Job	.66	.61	.76	.50
7. Discipline	.60	.87	.62	.58
8. Application and Initiative	.49	.59	.60	.54
9. Judgment and Common Sense	.61	.67	.69	.46
10. Dependability	.47	.43	-	-
11. Adaptability	.53	.63	-	-
12. Leadership	.60	.55	.72	.56
13. Overall Efficiency	.54	.62	.65	.51
14. Neatness of Work	.51	.48	.64	.44
15. Ability to be Taught	.42	.53	.55	.42
16. Care of Equipment	.42	.61	.55	.35
17. Ability to Troubleshoot	.60	.54	-	-
18. Sincerity in Job	.44	.60	.60	.44
19. Manual Skill	.58	.45	-	-
20. Overall Proficiency in Rate	.57	.63	.66	.40
Average Rating Scale Score	.67	.73	.74	.59

* Tetrachoric Correlation Coefficients

these traits are much easier to define and their manifestations easier to observe, which may explain the greater degree of agreement on them. Further, they are all traits which reflect a basic ability to perform and, as such, may tend to correlate with each other.

The factor of pay grade raised or lowered the inter-rater agreement depending upon circumstances. In the first sample, the removal of pay grade variance caused the majority of the coefficients to increase, while a few remained the same and a few decreased in size. In general, the increase in agreement occurred with traits which are not so highly related to pay grade. Thus it may have been that instructions designed to decrease inter-trait correlations and reduce halo were somewhat effective with the preliminary sample, and this fact showed up in the inter-rater agreement. It was apparent, in scoring the ratings, that pay grade was affecting various raters differently. Some were overcome by its influence and rated strictly in accordance with pay grade level on every trait. Others remained more objective, their ratings in many traits showing much smaller relationships to pay grade. If both kinds of raters picked the same men within pay grade level as best and poorest, then removal of the effect of pay grade should increase inter-rater agreement. It is believed that this is what happened in the first sample.

In the revised preliminary sample, however, the general effect of removing pay grade variance was to reduce the inter-rater agreement. It is concluded that the reason for this reversal of the earlier results was due to the instructions (this time the lack of them) given to the raters. Since there was practically no personal contact with the raters during the second administration of the rating scale, and since many of the raters in

the second sample had not participated in the first ratings, it is reasonable to hypothesize that the logical and halo effects created by the presence of pay grade were greater in the second sample than in the first. This hypothesis is substantiated by the fact that the correlations of pay grade with other rating scale variables was higher in the second sample than in the first, and also by the fact that trait intercorrelations were higher in general for the second sample. It is suggested, then, that pay grade variance contributed relatively more to the magnitude of the inter-rater agreement in the second sample, and its removal caused a reduction in agreement.

In concluding the discussion on the agreement of raters, a word is in order about the level of agreement obtained. In general it was more substantial, both on individual traits and on average score, than is usually reported with conventional scales. Undoubtedly exclusion of some of the less objective traits would have increased the agreement on average score appreciably.

Intercorrelations of traits and results of factorial analysis

A survey of the literature on rating scales revealed few studies which had been carried as far as factorial analysis.[1,2,3] These studies, with the exception of that by Bolanovich, plus well established indications of rating fallacies such as halo effect and logical errors leave one with the impression that seldom are more than one or two factors required to account for the high inter-trait correlations which typically are found in rating scale studies.

---

[1]Bolanovich, D. J., "Statistical analysis of an industrial relations chart," J. of Appl. Psychology, 30, 1946, 22-31.

[2]Chi, Pan-Lin, "Statistical analysis of personality ratings," J. of Experimental Education, 5, 1937, 229-245.

[3]Ewart, E., Seashore, S. E. and Tiffin, J., "A factor analysis of an industrial merit rating scale," J. of Appl. Psychology, 25, 1941, 401-486.

A part of the analysis of the present rating scale was to determine the magnitude of the inter-trait correlations and to perform factorial analyses of the resulting matrices.* This was done for each of two preliminary samples of submarine personnel, under three conditions for each sample with a resultant total of six analyses. The various conditions of these analyses are described below.

Intercorrelation of traits, Procedure I. In the initial analysis, traits were intercorrelated in a straightforward manner. Individuals were assigned plus scores in all traits in which they received average STEN-scores of five or better, and minus scores in all traits in which they received STEN-scores of four or less. Tetrachoric coefficients were then computed.

In addition to the rating scale traits, each man's general Classification Test Score (GCT), his age, length of time on board, education and pay grade level were included in the matrix for analysis.

In the Navy, pay grade is quite naturally related to most traits that make for success. This fact tends to increase the correlation between traits on a rating scale such as the one under discussion. As in any organization, those indiviudals who are at the most advanced levels are rated higher because of their positions and, in turn, they have been advanced to those levels because they possess more desirable amounts of the tra. that are necessary for advancement.

In view of this, and the relative naivete of the raters, one would expect the intercorrelations among the traits to be quite high. This

*Tables III through VIII, the intercorrelation matrices, will be found in Appendix A.

expectation was realized as the matrices in Table III (first sample) and Table IV (second sample) indicate. (Because of their size, these and the other correlational matrices, Tables V, VI, VII and VIII appear in Appendix A.) The intercorrelations in the second sample were somewhat higher than those obtained from the first sample. This was due probably to less personal instruction of the raters in the second sample, and to the greater dispersion of scores as a result of the instruction that no tie scores be given.

[A word is in order in regard to the reliability of the coefficients in these and subsequent calculations. The standard errors are less than the N's of 107 and 286 for the two samples would indicate. This is true because of the fact that the cell frequencies in the four-fold tables did not total N, but rather added up to the total number of pairs of ratings. Since there were two to three raters for every ratee, this total was considerably greater than N.]

Inspection of Tables III and IV indicates that pay grade is highly associated with technical skill as reflected in the traits of Knowledge of the Job, Judgment and Common Sense, Dependability, Leadership, Ability to Troubleshoot, and Age. Less highly related to pay grade is a group of traits which is considered to indicate the degree of personal adjustment to Navy life. In this group are Social Adjustment, Neatness of Appearance, Cooperation, Discipline, Application and Initiative, Ability to be Taught, and Neatness of Work. Length on Board is seen to have a small relationship to most traits, and Education and GCT have practically none. (Submarine personnel are pre-selected on GCT.)

Intercorrelation of traits. Procedure II (Variance within pay grade). Because of the substantial relationship of pay grade to most of the rating scale trait scores, and because of the great range of abilities represented in groups containing everyone from strikers to chief petty officers, within pay grade intercorrelations were computed next, treating a man's scores as high or low relative only to his own pay grade average. If some factor (s) other than pay grade was in part producing the intercorrelations, then the resulting matrices should contain significant coefficients. If, however, pay grade were the only variable operating to produce the intercorrelations, then the matrices would contain only near-zero coefficients.

In Tables V and VI it is possible to obtain a very general picture of the effect of removing pay grade variance. It will be noted that either method of removing pay grade reduced the intercorrelations of the traits appreciably, but did not by any means reduce them to near-zero. The amount of reduction can be discerned from the sums of the coefficients of the matrices. In the first sample, the sum of the coeffccients, Table III, was 162.39 (25 variables - disregarding the signs). In the reduced matrix, Table V (24 variables), it was 115.07. In the second sample, the sum of the coefficients, Table IV (21 variables), was 125.09. In the reduced matrix, Table VI (20 variables), the sum was 74.61 . This greater change in the second sample supports the hypothesis that lack of instructions during the second administration increased the halo due to pay grade, and thus its contribution to the total variance.

As testimony to the fact that the within pay grade variance technique

was effectively removing the variance due to pay grade level alone, the correlation of the rating scale variables with the trait of Age may be examined. In both samples, age correlated very highly with pay grade (.83 in the first sample, .94 in the second). Removing the effects of pay grade, then, should be tantamount to removing the effects of age, and the correlations of Age with the rating scale variables should distribute themselves rather closely about zero. Examinations of the correlations with age in Tables V and VI reveals that in both samples this did occur, over three-fourths of the coefficients having values between plus and minus .15. From this evidence it appears that the reducing procedure accomplished its purpose.

Intercorrelation of traits, Procedure III (pay grade partialled out). As an additional check on the technique of reducing within pay grade variance, however, it was decided to compare these results with those obtained by partialling out statistically the effects of pay grade. This was accomplished by starting with the original matrices and removing the pay grade variance from each coefficient by conventional partialling techniques.[4] The partialled matrices resulting may be seen in Tables VII and VIII for the first and second samples respectively.

In comparing specific corresponding coefficients in the reduced and partialled matrices, considerable variation can be seen. Apparently the two procedures did not produce precisely the same effects. However, a comparison of the sums of these matrices for both samples yields striking similarities. In the first sample, the sum of the reduced

[4] J. P. Guilford, Fundamental Statistics in Psychology and Education (New York: McGraw-Hill Book Company, 1942), pp. 268-271.

matrix, Table V, was 115.07 and that of the partialled matrix, Table VII, was 117.67. In the second sample, the sum of the reduced matrix, Table VI, was 74.61 while that of the partialled matrix, Table VIII, was 79.86. Thus the two methods removed approximately the same amount of variance in each sample, but they treated individual coefficients somewhat differently. The explanation for this must lie in the fact that what was accomplished precisely by mathematics was dependent on the correlation of pay grade with the other variables derived from the entire sample, while in the reducing procedure each individual case required a decision as to who was high and who was low within each pay grade. In other words, if the same amount of variance were due to pay grade in the case of each coefficient, then the partialling procedure should correspond very closely to the reducing procedure. It is believed that this was not the case in every instance and thus discrepancies between the matrices resulting from the two procedures were to be expected.

As a final comment on the partialling procedure, it should be pointed out that the correlations of Age with the rating scale variables again distributed themselves about zero as they did in the reducing procedure. Again, over 70 percent of the correlations with age were within the range of plus or minus .15. There were a few more large coefficients than with the reducing procedure, however, and a larger number of negative coefficients.

The discussion on the reducing and partialling procedures should not be concluded without recognizing that these techniques probably removed

too much variance. That is to say, if many of the qualities under consideration are genuinely related to pay grade, then removal of the effects of pay grade eliminated real, as well as spurious, variance. Evidence that this was the case occurred particularly during the analysis of the second sample of data. When pay grade was partialled from this matrix, a substantial number of negative correlations arose between the rating scale traits and such variables as Age, Length in Service, and scores in the General Classification Test. This, in turn, required a reflection of those variables during the factoring procedure, which led eventually to a few rather substantial negative factor loadings for these variables. These loadings cannot be interpreted in the usual sense, and are regarded as artifacts of the reducing and partialling procedures.

Results of the factor analyses. Having computed the inter-correlation of traits under three conditions for each of the two samples, six factor analyses were next performed using the Thurstone method. The purpose of these analyses was two-fold: (1) to determine the extent and nature of the factor structure of the rating scale under each of the three conditions of correlations, and (2) to determine how similar the factor structure was from one sample to the next.

Extraction of factors was continued as long as the cross-product of any pair of factor loadings exceeded the standard error of the zero-order coefficient between the corresponding pair of traits in the original matrix. For the sake of consistency and in order to introduce a minimum of assumptions, all rotations were orthogonal and the

criteria for rotation were positive manifold and simple structure. Rotations were accomplished according to a graphic method described by Zimmerman.[5]

In assembling the traits which comprised the rating scale, it was felt that there were temperamental components of performance aboard submarines which were of as much importance as technical knowledge and skills. Theoretically, at least, the temperamental components should be independent of the technical components, although for any given population they might be correlated. For this reason, orthogonal rotations were utilized which, in fact, gave very satisfactory solutions for the original data from both samples. For the sake of determining whether or not the factor structure was altered by the reducing and partialling procedures, orthogonal rotations were employed in those analyses also. Here the results were not nearly as satisfactory from the standpoint of identifying factors. It may be that an oblique solution would have resulted in more meaning for these sets of data. However, such an investigation was not considered advisable because of the labor required.

In Tables IX, X and XI (Appendix A), the results of the factorial analyses performed on the data from the first sample may be seen. In Tables XII, XIII and XIV, are the corresponding results from the second sample. The most remarkable feature of these results is the fact that the factor structure increased in dimensionality when variance due to pay grade was removed. In the first sample, there were four

[5]Zimmerman, W. S., "A simple graphical method for orthogonal rotation of axes", Psychometrika, 11:51-55, 1946.

factors before the effects of pay grade were removed, five factors in the reduced data, and six in the data in which pay grade was held constant by partial correlation. In the second sample this increase was from three to five and seven factors for the matrices from which pay grade was removed. No obvious explanation for this is readily available. One reasonable hypothesis might be, however, that in the first analysis for each sample, pay grade acted as a general factor and increased the correlation between factors to such an extent that two or more factors emerged as one.

Identification of factors. The most readily identifiable factors were extracted from the original matrices of both samples. In both these analyses a factor called Technical Competence emerged which was practically synonomous with pay grade. The traits and other variables having the highest loadings in these factors are listed below:

First Sample	Technical Competence
Pay Grade	.97
Knowledge of Job	.78
Age	.75
Leadership	.69
Judgment and Common Sense	.64
Ability to Troubleshoot	.63

Second Sample	Technical Competence
Pay Grade	.99
Age	.95
Length in Service	.92
Judgment and Common Sense	.80
Knowledge of Job	.80
Leadership	.78
Quality of Work	.74
Overall Efficiency in Rate	.73

In both original analyses a factor also emerged which seemed best identified as Personal Adjustment of some kind. Involved in it were attitude toward the job and shipmates, effort, sincerity, dependability, etc. The traits having the highest loadings in this factor for each sample are listed below:

First Sample	Personal Adjustment
Cooperation	.79
Application and Initiative	.77
Ability to be Taught	.75
Adaptability	.73
Sincerity in the Job	.69
Overall Efficiency	.69
Care of Equipment	.67
Dependability	.64

Second Sample	Personal Adjustment
Ability to be Taught	.75
Watch Standing	.70
Cooperation	.68
Application and Initiative	.65
Care of Equipment	.64
Sincerity in the Job	.64
Social Adjustment	.62
Quality of Work	.60
Overall Efficiency	.60

Also identified from the original data of the first sample was a factor which appears best described as Carefulness or Neatness in Work and Person. Traits having the highest loadings were:

First Sample	Carefulness or Neatness
Quality of Work	.56
Neatness of Work	.52
Care of Equipment	.50
Overall Efficiency in Rate	.50
Discipline	.45
Neatness of Appearance	.40

This rather well defined factor was not identified in the analysis of the data from the second sample.

The fourth and final factor emerging from the original data of the first sample appeared to be some sort of Efficiency or Job Performance factor. The traits with principal loadings were:

First Sample	Efficiency or Job Performance
Neatness of Work	.59
Overall Efficiency in Rate	.57
Watch Standing	.55
Overall Efficiency	.46
Sincerity in the Job	.45
Manual Skill	.44
Ability to Troubleshoot	.37
Judgment and Common Sense	.37
Quality of Work	.34

The third and final factor emerging from the original data of the second sample was found only in that sample. While all loadings were low, it appeared to be best identified with Maturity or Experience:

Second Sample	Maturity or Experience
Length on Board	.46
Education	.41
Knowledge of Job	.30
Judgment	.35
Leadership	.33

The factors extracted from the reduced and partialled matrices were very difficult to identify. Traces of the factors extracted from the original data, listed above, could be seen throughout. The loadings, of course, were very different due to the removal of pay grade variance. An attempt was made to classify the factors into two broad groups -- those of a Technical nature and those of an Adjustment nature. Even

with such a broad classification, some factors seemed to be as much a member of one class as of the other. Because of the possibly limited significance of these patterns of loadings, they will not be reported here.

The difficulty in identifying these last groups of factors probably rests partly with the fact that too few experimental variables were included for identification of the large number of factors, which theoretically might be rated. To this difficulty may be added the fact that the rating scale traits were not sufficiently definitive to permit a distinction between two factors which were comprised in large part of the same traits. For example, in the field of aptitudes, identification of a factor in which many tests have loadings can be made with considerable confidence if a relatively pure test (such as number operations or vocabulary) is highly saturated with the factor. In the present analysis, however, one would hesitate to identify a factor as Cooperation, for example, simply because the trait called Cooperation had the highest loading in that factor.

Identification of factors extracted from the reduced and partialled matrices was increased in difficulty also by the fact that these procedures removed a good proportion of genuine as well as spurious variance. This left many of the traits with reduced loadings in certain factors, and decreased differences in loadings between traits. This resulted in less distinct patterns of loadings which were correspondingly more difficult to identify.

Chapter IV

RESULTS OF THE SHIPBOARD USE OF THE RATING SCALE

B. Relationships with other Performance Measures

The performance rating scale was the first of three criterion measures developed in the study of shipboard performance of Navy enlisted personnel. The results reported in previous chapters indicate that it provides a means for obtaining reliable, discriminating measures of at least two general aspects of shipboard performance -- one technical and one adjustive in nature.

Later in the study, performance check lists were developed to determine how well officers and petty officers could rate their men on ability to perform specific tasks from their jobs.[1] These were much more specific and technical in nature than the general traits of the rating scale.

Finally, actual performance tests were constructed around tasks representative of the EM and EN jobs aboard submarines.[2] These provided an opportunity to compare rated performance with actual tested performance.

Correlation of rating scale and check list evaluations

It would be expected that scores from the two rating measures would correlate appreciably, simply because scores on both were substantially related to pay grade. Part of the relationship of pay grade to the ratings reflected the genuinely higher average performance level of men in the higher pay grades, but part, no doubt also reflected the tendency of some raters to rate men high simply because they were of higher pay grade and "should" be able to perform better. It was decided, therefore, that the relationships

[1] See Part III, this report.

[2] See Part I, this report.

between the two measures should be investigated on a within pay grade basis. In order to have an estimate of the highest likely relationship, however, it was decided to correlate, across all pay grades, scores from the check lists with perhaps the most technical rating scale trait, Ability to Troubleshoot.

The results, broken down by officer and enlisted rater, and by observed and total check list scores,[1] appear in Table XV. The very high relationships obtained, particularly in the case of check list total scores, suggest that the check list and technical traits of the rating scale were measuring essentially the same things if scorss were taken across pay grade.

TABLE XV

CORRELATIONS BETWEEN SCORES ON THE RATING SCALE TRAIT "ABILITY TO TROUBLESHOOT" AND VARIOUS CHECK LIST SCORES (ALL PAY GRADES INCLUDED)

SCORE	EM	EN
Enlisted rater, observed scores	.78 (N = 229)	.71 (N = 246)
Officer rater, observed scores	.53 (N = 100)	.52 (N = 142)
Enlisted rater, total scores	.89 (N = 234)	.83 (N = 293)
Officer rater, total scores	.90 (N = 159)	.79 (N = 215)

The within pay grade study was made on striker and third-class petty officers only. These were the groups singled out for special study as a result of the overall design of the project. (As a part of this study, specially designed aptitude tests had been administered to men entering the Submarine School, New London. It was intended that the relationships between these tests and the shipboard criterion measures would be examined. Therefore attention was focused on strikers and thirds since few higher rated petty officers would have been recent Submarine School graduates.)

[1] See Part III, this report for a discussion of "observed" check list scores.

The results appear in Table XVI. Correlations are reported for strikers and thirds separately and also across pay grade to study the influence of that variable at the lower pay grade levels.

TABLE XVI

RELATIONSHIPS BETWEEN CHECK LIST AND RATING SCALE TOTAL SCORES

(Striker and Third-Class Petty Officers, Biserial Correlations)

	EM			EN		
	Strikers	Thirds	Strikers & Thirds	Strikers	Thirds	Strikers & Thirds
Enlisted rater, observed scores	.65	.53	.58	.20	.42	.50
Officer rater, observed scores	.13	.40	.37	.25	.40	.44
Enlisted rater, total scores	.46	.45	.61	.41	.61	.56
Officer rater, total scores	.58	.58	.54	.38	.23	.40

The last two lines in Table XVI reveal the correlations between total scores on the check lists and total rating scale scores to be substantial even within the lower pay grades and in spite of the fact that some of the rating scale variance was non-technical in nature. The inclusion of pay grade variance does not appear to increase these relationships systematically at these job levels.

As a final check on the relationships between check list and rating scale, scores on each of the twenty individual traits on the preliminary rating scale were correlated with total check list scores. These results appear in Table XVII. It can be seen that, even on a within pay grade basis, the more technically oriented rating scale traits tended to correlate higher with the check list scores than those which were more adjustive or attitudinal in nature.

TABLE XVII

RELATIONSHIPS BETWEEN PERFORMANCE CHECK LIST TOTAL SCORES AND SCORES ON INDIVIDUAL TRAITS OF THE PRELIMINARY RATING SCALE (Within Pay Grade N = 187)

Rating Scale Traits	Correlation (Tetrachoric) with Check List Scores
1. Social Adjustment	-.10
2. Quality of Work	.32**
3. Neatness of Appearance	.07
4. Cooperation	.14
5. Watch Standing	.32**
6. Knowledge of Job	.49**
7. Discipline	.29*
8. Application and Initiative	.16
9. Judgment and Common Sense	.40**
10. Dependability	.40**
11. Adaptability	.37**
12. Leadership	.35**
13. Overall Efficiency	.46**
14. Neatness of Work	.40**
15. Ability to be Taught	.21
16. Care of Equipment	.35**
17. Ability to Troubleshoot	.41**
18. Sincerity in the Job	.43**
19. Manual Skill	.43**
20. Overall Efficiency in Rate	.31**

* Significant at the 5% level.
** Significant at the 1% level.

For example, Knowledge of the Job, Judgment and Common Sense, Overall Efficiency, Ability to Troubleshoot and Manual Skill all correlated very significantly with total check list scores on a within pay grade basis. However, certain traits associated with the adjustment factor in the rating scale, including Social Adjustment, Neatness of Appearance, Cooperation, Application and Initiative, and Ability to be Taught did not correlate significantly.

Correlations between rating scale scores and performance test scores

As a final study of the nature of rating scale scores as criteria of shipboard performance, correlations were computed with scores on the job sample tests and written job knowledge examinations. Again, these relationships were studied first as they existed for all pay grades and later as they existed within the striker and third class petty officer levels.

Table XVIII reveals the across pay grade relationships between the ten traits on the final version of the rating scale and average job sample test scores and written job knowledge test scores. Generally, the traits Ability to Troubleshoot, Knowledge of the Job, Judgment and Common Sense, Leadership and Care of Equipment were more highly related to performance test scores than were other traits. This tendency was more pronounced for EM's than for EN's.

TABLE XVIII

CORRELATION BETWEEN ACROSS-PAY GRADE RATING SCALE TRAIT SCORES, AVERAGE JOB SAMPLE TEST SCORES, AND JOB KNOWLEDGE TEST SCORES
(Tetrachoric Coefficients)

	EM's*		EN's*	
Characteristic	Average Job Sample Score	Job Knowledge Score	Average Job Sample Score	Job Knowledge Score
Cooperation	.25	.36	.38	.40
Knowledge of Job	.50	.48	.34	.51
Discipline	.30	.32	.36	.35
Application and Initiative	.41	.38	.32	.40
Judgment and Common Sense	.53	.47	.38	.41
Leadership	.57	.52	.45	.44
Neatness of Work	.43	.44	.46	.45
Care of Equipment	.53	.38	.47	.36
Ability to Troubleshoot	.62	.53	.55	.56
Sincerity in Job	.43	.36	.44	.47

* For EM's, N = 315 Strikers through CPO's; for EN's, N = 403 Strikers through CPO's.

A similar analysis next was done on a within pay grade basis, including strikers and third class petty officers only. These results appear in Table XIX. It will be noticed that the relationships are smaller in every case. The correlations still are appreciable for the EM's but in the case of EN's, the correlations of job sample test scores with the rating scale traits reduced to nearly zero. This was not true of the correlations between rating traits and the written job knowledge test, however. For the written test the drop in correlations was quite similar for both EM's and EN's.

TABLE XIX

CORRELATIONS BETWEEN WITHIN-PAY GRADE RATING SCALE TRAIT SCORES, AVERAGE JOB-SAMPLE TEST SCORES, AND WRITTEN JOB-KNOWLEDGE TEST SCORES
(Tetrachoric Coefficients)

	EM's*		EN's*	
Characteristic	Average Job Sample Score	Job Knowledge Score	Average Job Sample Score	Job Knowledge Score
Cooperation	.20	.20	.02	.27
Knowledge of Job	.48	.25	.16	.32
Discipline	.28	-.16	.04	.10
Application and Initiative	.30	.20	.04	.34
Judgment and Common Sense	.42	.36	.13	.30
Leadership	.38	.27	.04	.25
Neatness of Work	.45	.27	.10	.28
Care of Equipment	.38	.23	.10	.35
Ability to Troubleshoot	.39	.32	.23	.36
Sincerity in Job	.40	.21	.23	.39

* For EM's, N = 147 Strikers and 3rd Class PO's; for EN's, N = 190 Strikers and 3rd Class PO's.

In addition to the correlations for the individual traits of the rating scale, total scores also were correlated with average performance test scores. These values proved to be .14 for EN's and .38 for EM's. Thus, as might be expected, certain of the more technically oriented rating scale traits correlated higher with tested performance than did the average of scores from all traits.

Conclusion

Some of the results of the foregoing analyses may be regarded as encouraging and some discouraging as far as the relationship between rated and tested performance is concerned.

Relationships between the two kinds of measures tend to be substantial if considered across pay grade. This would be true, however, because both are, themselves, related to pay grade.[1]

Within pay grade the correlation was appreciable for EM's but not so for EN's. It would be highly desirable if rated performance would accurately reflect tested performance within each pay grade level. However, this is demanding reliable and valid observation of technical performance within very narrow ranges of skill and job experience. Probably most Navy raters have not gathered sufficient specific evidence from their incidental observations of men on the job to know the precise nature and rank order of skills among men in the lower pay grades.[2]

It is also quite possible that the performance tasks used in this study were not sufficiently comprehensive to reflect the variety and depth of skills demanded in the two rates studied. This probably was more true of the EN than of the EM battery and the correlations throughout the study lend credence to this notion.

Finally, it must be remembered that the rating scale was designed to reflect aspects of shipboard performance other than the technical ones. Job attitude or adjustment to Navy life must be considered highly important, particularly in the Submarine Navy. These aspects of performance are possibly greater determiners of

---

[1] The correlation between pay grade and performance test scores was .36. That between pay grade and ratings was .62 in the preliminary and .69 in the revised preliminary scale.

[2] This conclusion also is borne out by the results reported in Part III of this report.

the ratings a man receives from his superiors than are technical knowledge and skill.

It would seem, then, that reliable and discriminating ratings by superiors would form a necessary part of the criterion of any man's shipboard performance. However, the within-pay-grade correlations between ratings and tests of performance in this study were so low that it must be considered desirable, in addition, to actually test a man on his practical factors wherever possible.

Chapter V

CONCLUSIONS

The use by Navy personnel of a rating scale designed to measure the shipboard performance of submarine personnel has led to the following conclusions:

1. Reliable and discriminating evaluations of performance aboard ship can be obtained with a rating scale that is properly designed and used. The "man-to-man" format adopted for this study appears to be practical and tends to promote better ratings than conventional scales.
2. Reliable ratings can be obtained on single traits as well as on total scores; reasonably reliable ratings can be obtained within-pay-grade, even at the striker level.
3. Navy raters need specific instructions on the following if their ratings are to be highly useful:
    a) halo effect, particularly that due to pay grade;
    b) the logical error leading to high intercorrelation of traits;
    c) the need for maximum discrimination in ratings;
    d) the need for a realistic average in the ratings assigned.

    To a certain extent written instructions accomplish the needed education of raters. Personal instruction is to be preferred, however.
4. At least two broad aspects of shipboard performance can be tapped by means of ratings. In this study these appeared to be technical competence or job skill on the one hand, and personal adjustment or attitude toward the job on the other.
5. Appraisals of performance made with the rating scale correlate with

actual tests of performance to a very limited extent if the range of experience is restricted to a single pay grade. This is apparently more true for some rates and pay grades than others.

6. Inter-rater differences in leniency must be accounted for if high agreement and reliability are to be realized. This is particularly true if such agreement is sought on a within pay grade basis.

The usefulness of a rating device depends on how it is designed and how well it is used. The device should include well defined traits and performance levels. It should be designed so that men may be readily compared with one another within the framework of a single performance characteristic. It should promote discriminating ratings. Ideally, the raters should be aware of the common pitfalls of rating and the many uses for good ratings. Under these conditions, highly useful evaluations of performance can be made.

Because of the limited correlations between ratings and performance test scores obtained in this study, it must be concluded further, that a satisfactory criterion of shipboard performance demands the inclusion both of ratings by superiors and actual tests of performance.

SELECTED BIBLIOGRAPHY

BIBLIOGRAPHY*

A. BOOKS

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1942. 333 pp.

_____, Psychometric Methods. New York: McGraw-Hill Book Company, 1936. 566 pp.

Mahler, W. R., Twenty Years of Merit Rating (1926-1946). New York: The Psychological Corporation, 1947. 73pp.

B. PERIODICAL ARTICLES

Almy, H. C., and H. Sorenson, "A Teacher Rating Scale of Determined Reliability and Validity," Educational Administration and Supervision, 16, 1930, 179-186.

Barteau, Charles E., "A New Conception in Personnel Rating," Personnel, 13, 1936, 20-27.

Bittner, R. H., "Developing an Industrial Merit Rating Procedure," Personnel Psychology, 1, 1948, 403-432.

Bittner, R. H., and E. A. Rundquist, "The Rank-Comparison Rating Method," Journal of Applied Psychology, 34, 1950, 171-177.

Blum, M. L., "A Contribution to Manual Aptitude Measurement in Industry," Journal of Applied Psychology, 24, 1940, 381-416.

Bolanovich, D. J., "Statistical Analysis of an industrial Relations Chart," Journal of Applied Psychology, 30, 1946, 22-31.

Bradshaw, F. F., "The American Council on Education Rating Scale: Its Reliability, Validity and Use," Archives of Psychology, 119, 1930, 80 pp.

---

* This bibliography does not include references after 1950 since construction of the rating scale was completed at that time.

Champney, H., and H. Marshall, "Optimal Refinement of the Rating Scale," Journal of Applied Psychology, 23, 1939, 323-331.

Chi, Pan-Lin, "Statistical Analysis of Personality Ratings," Journal of Experimental Education, 5, 1937, 229-245.

Davis, William S., "Factor Merit Rating System," Personnel, 22, 1946, 309-319.

Driver, R. S., "The Validity and Reliability of Ratings," Personnel, 17, 1941, 185-191.

Evans, J. W., "Emotional Bias in Merit Rating," Personnel Journal, 28, 1950, 290-291.

_____, "The Rater's Task in Merit Rating," Personnel Journal, 28, 1950, 375.

Ewart, E., S. E. Seashore, and J. Tiffin, "A Factor Analysis of an Industrial Merit Rating Scale," Journal of Applied Psychology, 25, 1941, 481-486.

Ferguson, L. W., "The Value of Acquaintance Ratings in Criterion Research," Personnel Psychology, 2, 1949, 93-102.

_____, "A Brief Description of a Reliable Criterion of Job Performance," Journal of Psychology, 25, 1948, 389-399.

Flannagan, J. C., "A New Approach to Evaluating Personnel," Personnel, 26, 35-42, 1949.

_____, "Critical Requirements: A New Approach to Employee Evaluation," Personnel Psychology, 2, 1949, 419-425.

Furfey, P. H., "An Improved Rating Scale Technique," Journal of Educational Psychology, 17, 1926, 45-48.

Gallup, G. H., "Traits of Successful Retail Salespeople," Journal of Personnel Research, 4, 1926, 474-482.

Garrison, K. C., and S. C. Howell, "The Relationship Between Character Trait Ratings and Certain Mental Abilities," Journal of Applied Psychology, 15, 1931, 378-389.

George, Wally E., "How We Rate Apprentices," Factory Management and Maintenance, 95, 1937, 55pp.

Gilinski, A. S., "The Influence of the Procedure of Judging on the Halo Effect," American Psychologist, 2, 1947, 309-310.

Jurgensen, C. E., "A Fallacy in the Use of Median Scale Values in Employee Check Lists," Journal of Applied Psychology, 33, 1949, 56-59.

Kidder, J. H. T., "Employee Rating Methods of Appraising Ability, Efficiency, and Potentialities," National Industrial Conference Board Management Record, 4, 1942, 33-40.

King, J. E. "Multiple-Item Approach to Merit Rating," American Psychologist, 4, 1949, 278 pp. (Abstract).

Hausman, H. J., J. T. Begley, and H. L. Parris, "Selected Measures of Proficiency for B-29 Mechanics: Study No. 1," Human Resources Research Laboratories Report #7. July, 1949.

Knauft, E. B., "A Classification and Evaluation of Personnel Rating Methods," Journal of Applied Psychology, 31, 1947, 617-625.

_____, "Construction and Use of Weighted Check List Rating Scales for Two Industrial Situations," Journal of Applied Psychology, 32, 1948, 63-70.

Kornhauser, A. W., "Reliability of Average Ratings," Journal of Personnel Research, 5, 1926, 309-317.

Lawshe, C. H., N. C. Kephart, and E. J. McCormick, "The Paired-Comparison Technique for Rating Performance of Industrial Employees," Journal of Applied Psychology, 33, 1949, 69-77

Mahler, W. R., "Some Common Errors in Employee Rating Practices," Personnel Journal, 26, 1947, 68-74.

_____, "An Experimental Study of Two Methods of Rating Employees," Personnel, 25, 1948, 211-220.

Marble, S. D., "A Performance Basis for Employee Evaluation," Personnel, 18, 1942, 217-226.

Markey, S. C., "Consistency of Descriptive Personality Phrases in the Forced-Choice Technique," American Psychologist, 2, 1947, 310-311.

Marsh, S., and F. Perrin, "An Experimental Study of Rating Scale Technique," Journal of Abnormal and Social Psychology, 19, 1925, 383-399.

Morrow, L., "A New Approach to Merit Rating," Modern Management, 9, 1949, 19-22.

Patterson, C. H., "On the Problem of the Criterion in Prediction Studies," Journal of Consulting Psychology, 10, 1946, 277-280.

Pockrass, J. H., "Rating Training and Experience in Merit System Selection," Public Personnel Review, 2, 1941, 211-222.

Richardson, M. W., "Forced-Choice Performance Reports," Personnel, 26, 1949, 205-212.

_____, "An Experimental Study of the Forced-Choice Performance Report," American Psychologist, 4, 1949, 278-279. (Abstract).

Robertson, A. E., and C. L. Stromberg, "The Agreement Between Associates' Ratings and Self-Ratings of Personality," School and Society, 50, 1939, 126-127.

Ryan, T. A., "Merit Rating Criticized," Personnel Journal, 24, 1945, 6-15.

Sisson, E. D., "Forced-Choice--The New Army Rating," Personnel Psychology, 1, 1948, 365-381.

Smith, I. S., "Developing a Service Rating System," Educational and Psychological Measurement, 4, 1944, 327-337.

Stevens, S. N., and E. F. Wonderlic, "An Effective Revision of the Rating Technique," Personnel Journal, 13, 1934, 125-134.

Stockford, L., and H. W. Bissell, "Factors Involved in Establishing a Merit Rating Scale," Personnel, 26, 1949, 94-116.

Taylor, E. K., "What Raters Rate," American Psychologist, 3, 1948, 289-290. (Abstract).

Tiffin, J., and W. Musser, "Weighting Merit Rating Items," Journal of Applied Psychology, 26, 1942, 575-583.

Turner, W. D., "A Multiple Committee Method of Merit Rating," Personnel, 25, 1948, 176-194.

Viteles, M. S., "A Psychologist Looks at Job Evaluation," Personnel, 17, 1941, 165-176.

Wadsworth Jr., G. W., "The Field Review Method of Employee Valuation and Internal Placement," Personnel Journal, 27, 1948, 47-54.

Weinstock, I., "Merit Rating--A Restatement of Principles," Personnel Journal, 27, 1948, 223-226.

Weiss, R. A., "Rating Scales," Psychological Bulletin, 30, 1933, 185-208.

Wherry, R. H., and D. H. Fryer, "Buddy Ratings: Popularity Contest or Leadership Criteria?" Personnel Psychology, 2, 1949, 147-159.

Wiebe, G. D., "A Comparison of Various Rating Scales Used in Judging the Merits of Popular Songs," Journal of Applied Psychology, 23, 1939, 18-22.

Wilke, W. H., "The Reliability of Summaries of Rating Scale Evaluations of Student Personality Traits," Journal of Genetic Psychology, 53, 1938, 313-320.

Zerga, J. E., "Developing an Industrial Merit Rating Scale," Journal of Applied Psychology, 27, 1943, 190-195.

Zimmerman, W. S., "A Simple Graphical Method for Orthogonal Rotation of Axes," Psychometrika, 11, 1946, 51-55.

APPENDIX A

TABLE III

INTERCORRELATIONS* OF RATING SCALE TRAITS, NAVY
G.C.T. AND BIOGRAPHICAL INFORMATION ACROSS PAY GRADE (FIRST SAMPLE)
N = 187 Enginemen and Electrician's Mates

Trait	1	2	3	4	5	6	7	8	9	10	11	12
1. Pay Grade	-											
2. Age	.83	-										
3. Length on Board	.43	.23	-									
4. Education	-.12	.00	-.27	-								
5. Social Adjustment	.19	.08	.04	-.03	-							
6. Quality of Work	.57	.48	.16	.06	.44	-						
7. Neatness of Appearance	.41	.30	.13	-.05	.15	.42	-					
8. Cooperation	.34	.31	.14	-.13	.61	.68	.33	-				
9. Watch Standing	.71	.48	.29	-.14	.51	.74	.49	.66	-			
10. Knowledge of Job	.86	.68	.40	.01	.35	.77	.35	.47	.76	-		
11. Discipline	.38	.34	.20	.21	.33	.51	.47	.45	.54	.38	-	
12. Application and Initiative	.33	.29	.22	.01	.55	.75	.32	.84	.65	.59	.52	-
13. Judgment and Common Sense	.82	.68	.31	.12	.44	.81	.45	.58	.80	.93	.53	.62
14. Dependability	.80	.64	.25	.01	.47	.85	.53	.63	.82	.86	.53	.71
15. Adaptability	.64	.49	.36	.06	.57	.77	.40	.65	.81	.79	.43	.66
16. Leadership	.82	.72	.35	.04	.48	.77	.43	.56	.75	.88	.55	.62
17. Overall Efficiency	.56	.44	.30	.05	.62	.85	.49	.70	.82	.75	.52	.71
18. Neatness of Work	.53	.46	.15	.08	.46	.83	.44	.65	.72	.66	.37	.73
19. Ability to be Taught	.47	.46	.47	-.08	.42	.79	.41	.69	.74	.68	.55	.77
20. Care of Equipment	.60	.52	.29	.01	.42	.79	.56	.64	.83	.74	.68	.68
21. Ability to Troubleshoot	.82	.57	.43	-.10	.42	.81	.48	.55	.72	.86	.45	.71
22. Sincerity in Job	.62	.42	.34	-.04	.53	.83	.39	.74	.82	.77	.56	.89
23. Manual Skill	.71	.51	.27	.00	.43	.74	.35	.54	.62	.77	.41	.55
24. Overall Efficiency in Rate	.63	.57	.25	.11	.49	.79	.49	.70	.72	.77	.55	.83
25. Gen. Classification Test	.34	.16	-.01	.07	.13	-.02	.14	.04	.16	.27	.10	.11

* Tetrachoric Correlation Coefficients

TABLE III (Continued)

Trait	13	14	15	16	17	18	19	20	21	22	23	24	25
13. Judgment and Common Sense	-												
14. Dependability	.88	-											
15. Adaptability	.83	.80	-										
16. Leadership	.88	.87	.79	-									
17. Overall Efficiency	.82	.85	.89	.77	-								
18. Neatness of Work	.78	.81	.81	.68	.81	-							
19. Ability to be Taught	.74	.79	.87	.64	.84	.82	-						
20. Care of Equipment	.81	.78	.69	.77	.82	.74	.75	-					
21. Ability to Troubleshoot	.84	.81	.86	.85	.82	.76	.79	.71	-				
22. Sincerity in Job	.83	.85	.83	.79	.89	.81	.76	.81	.86	-			
23. Manual Skill	.78	.80	.76	.74	.75	.83	.76	.65	.80	.78	-		
24. Overall Efficiency in Rate	.85	.84	.88	.78	.85	.89	.85	.74	.86	.88	.77	-	
25. Gen. Classification Test	.17	.27	.15	.17	.21	-.01	.04	.20	.08	.08	.22	.08	-

TABLE IV

INTERCORRELATIONS* OF RATING SCALE TRAITS, NAVY G.C.T. AND BIOGRAPHICAL INFORMATION ACROSS PAY GRADE (SECOND SAMPLE)
N = 286 Enginemen and Electrician's Mates

Trait	1	2	3	4	5	6	7	8	9	10
1. Pay Grade	-									
2. Age	.94	-								
3. Length on Board	.12	.11	-							
4. Education	-.06	.00	-.27	-						
5. Social Adjustment	.46	.36	.08	-.01	-					
6. Quality of Work	.76	.68	.21	-.01	.67	-				
7. Cooperation	.65	.57	.15	.07	.74	.84	-			
8. Watch Standing	.63	.60	.18	.05	.68	.88	.82	-		
9. Knowledge of Job	.83	.78	.23	.04	.64	.91	.79	.83	-	
10. Discipline	.47	.37	.04	.16	.47	.67	.62	.65	.57	-
11. Application and Initiative	.59	.58	.21	.05	.67	.87	.80	.82	.80	.57
12. Judgment and Common Sense	.84	.80	.18	-.11	.65	.91	.78	.85	.93	.62
13. Leadership	.83	.72	.16	-.01	.68	.92	.82	.82	.95	.56
14. Neatness of Work	.74	.68	.17	.06	.67	.90	.77	.86	.87	.54
15. Ability to be Taught	.51	.51	.05	.14	.63	.84	.82	.83	.71	.65
16. Care of Equipment	.69	.68	.21	-.01	.69	.88	.82	.86	.88	.64
17. Sincerity in Job	.70	.67	.13	.02	.70	.89	.83	.85	.86	.62
18. Overall Efficiency in Rate	.66	.69	.20	.02	.58	.86	.73	.84	.87	.53
19. Overall Efficiency	.69	.61	.21	-.01	.67	.88	.80	.85	.89	.63
20. Length in Service	.95	.96	.07	-.09	.34	.70	.57	.55	.80	.27
21. Gen. Classification Test	.30	.19	-.07	.24	.11	.18	.12	.21	.18	.21

* Tetrachoric Correlation Coefficients

TABLE IV (Continued)

Trait	11	12	13	14	15	16	17	18	19	20	21
11. Application and Initiative	-										
12. Judgment and Common Sense	.77	-									
13. Leadership	.75	.92	-								
14. Neatness of Work	.81	.84	.90	-							
15. Ability to be Taught	.76	.80	.75	.77	-						
16. Care of Equipment	.84	.84	.83	.87	.82	-					
17. Sincerity in Job	.80	.83	.85	.86	.76	.88	-				
18. Overall Efficiency in Rate	.74	.88	.88	.79	.80	.81	.82	-			
19. Overall Efficiency	.74	.87	.90	.81	.78	.86	.85	.94	-		
20. Length in Service	.54	.79	.76	.71	.48	.66	.67	.62	.61	-	
21. Gen. Classification Test	.13	.19	.24	.20	.27	.11	.12	.23	.19	.14	-

TABLE V

INTERCORRELATIONS* OF RATING SCALE TRAITS, NAVY G.C.T.
AND BIOGRAPHICAL INFORMATION BASED ON VARIANCE WITHIN PAY GRADE (FIRST SAMPLE)
N = 187 Enginemen and Electrician's Mates

Trait	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	-											
2. Length on Board	-.02	-										
3. Education	.03	-.21	-									
4. Social Adjustment	-.01	.02	-.01	-								
5. Quality of Work	.23	.01	.24	.55	-							
6. Neatness of Appearance	-.07	-.04	.19	.09	.40	-						
7. Cooperation	.16	-.04	-.08	.56	.55	.38	-					
8. Watch Standing	.00	-.03	.04	.30	.53	.40	.46	-				
9. Knowledge of Job	.17	.12	.02	.53	.63	.38	.59	.65	-			
10. Discipline	-.16	.11	.14	.17	.54	.57	.20	.32	.33	-		
11. Application and Initiative	.02	-.02	.08	.49	.62	.37	.70	.62	.55	.36	-	
12. Judgment and Common Sense	.10	.02	.15	.25	.55	.50	.59	.63	.74	.36	.61	-
13. Dependability	.25	.13	.09	.39	.70	.45	.62	.68	.79	.61	.63	.65
14. Adaptability	.03	.06	.05	.41	.52	.44	.64	.57	.71	.23	.55	.64
15. Leadership	.12	.01	.14	.33	.51	.52	.43	.67	.68	.22	.59	.63
16. Overall Efficiency	.09	.13	.09	.53	.77	.53	.67	.65	.78	.43	.66	.69
17. Neatness of Work	.09	-.11	-.01	.19	.73	.38	.38	.68	.56	.37	.52	.50
18. Ability to be Taught	.00	.06	-.14	.48	.58	.36	.59	.46	.57	.34	.58	.42
19. Care of Equipment	.12	-.03	.07	.37	.77	.49	.57	.59	.65	.56	.62	.60
20. Ability to Troubleshoot	.16	.16	-.04	.48	.68	.39	.51	.61	.66	.15	.58	.63
21. Sincerity in Job	-.04	-.08	.12	.41	.68	.47	.61	.70	.69	.49	.69	.59
22. Manual Skill	-.01	.09	.09	.50	.63	.29	.50	.55	.53	.33	.43	.41
23. Overall Efficiency in Rate	.05	.03	.23	.42	.58	.39	.56	.46	.64	.27	.66	.69
24. Gen. Classification Test	-.01	-.08	.07	.13	.18	.03	.23	.21	.13	.12	.24	.23

* Tetrachoric Correlation Coefficients

TABLE V (Continued)

Trait	13	14	15	16	17	18	19	20	21	22	23	24
13. Dependability	-											
14. Adaptability	.62	-										
15. Leadership	.68	.68	-									
16. Overall Efficiency	.81	.80	.68	-								
17. Neatness of Work	.62	.50	.52	.72	-							
18. Ability to be Taught	.50	.62	.37	.70	.64	-						
19. Care of Equipment	.71	.63	.59	.73	.71	.53	-					
20. Ability to Troubleshoot	.72	.75	.69	.85	.72	.67	.64	-				
21. Sincerity in Job	.72	.71	.60	.80	.67	.57	.68	.68	-			
22. Manual Skill	.54	.67	.46	.61	.72	.56	.52	.60	.67	-		
23. Overall Efficiency in Rate	.59	.83	.63	.86	.59	.56	.58	.79	.74	.51.	-	
24. Gen. Classification Test	.16	.23	.18	.16	-.07	.06	.12	.05	.20	.23	.19	-

TABLE VI

INTERCORRELATIONS* OF RATING SCALE TRAITS, NAVY G.C.T. AND BIOGRAPHICAL INFORMATION BASED ON VARIANCE WITHIN PAY GRADE (SECOND SAMPLE)
N = 286 Enginemen and Electrician's Mates

Trait	1	2	3	4	5	6	7	8	9	10
1. Age	-									
2. Length on Board	-.06	-								
3. Education	.04	-.19	-							
4. Social Adjustment	.14	.23	.02	-						
5. Quality of Work	.05	.14	.30	.53	-					
6. Cooperation	.08	.07	.15	.74	.65	-				
7. Watch Standing	-.03	.25	.07	.47	.64	.46	-			
8. Knowledge of Job	.24	.21	.18	.49	.72	.56	.61	-		
9. Discipline	-.06	.07	.22	.26	.45	.47	.37	.37	-	
10. Application and Initiative	.13	.17	.25	.52	.74	.70	.54	.56	.45	-
11. Judgment	.33	.18	.18	.55	.70	.64	.59	.72	.44	.60
12. Leadership	-.06	.11	.14	.47	.73	.58	.59	.71	.44	.56
13. Neatness of Work	.00	.05	.23	.50	.69	.55	.58	.53	.43	.63
14. Ability to be Taught	-.06	.10	.22	.52	.74	.63	.70	.52	.45	.65
15. Care of Equipment	.08	.12	.22	.37	.74	.61	.50	.54	.51	.72
16. Sincerity in Job	.12	.19	.18	.53	.72	.63	.60	.61	.43	.75
17. Overall Efficiency in Rate	.08	.24	.20	.50	.83	.57	.69	.67	.51	.61
18. Overall Efficiency	-.06	.10	.14	.62	.78	.65	.75	.70	.45	.67
19. Length in Service	.45	.01	-.14	-.03	-.21	-.14	-.06	.14	-.19	-.12
20. Gen. Classification Test	.02	.04	.28	.25	.13	.07	.06	.07	.09	.05

* Tetrachoric Correlation Coefficients

TABLE VI (Continued)

Trait	11	12	13	14	15	16	17	18	19	20
11. Judgment	-									
12. Leadership	.63	-								
13. Neatness of Work	.58	.58	-							
14. Ability to be Taught	.62	.64	.63	-						
15. Care of Equipment	.56	.58	.59	.61	-					
16. Sincerity in Job	.64	.64	.74	.66	.69	-				
17. Overall Efficiency in Rate	.71	.71	.69	.74	.70	.74	-			
18. Overall Efficiency	.76	.77	.75	.73	.65	.72	.82	-		
19. Length in Service	.04	-.15	-.07	-.09	-.04	.10	-.07	-.11	-	
20. Gen. Classification Test	.20	.01	.05	.16	.14	-.05	-.05	.14	-.00	-

TABLE VII

INTERCORRELATIONS* OF RATING SCALE TRAITS, NAVY G.C.T. AND BIOGRAPHICAL INFORMATION, PAY GRADE PARTIALLED OUT STATISTICALLY (FIRST SAMPLE)
N = 107 Enginemen and Electrician's Mates

Trait	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	-											
2. Length on Board	-.25	-										
3. Education	.18	-.24	-									
4. Social Adjustment	-.14	-.05	.00	-								
5. Quality of Work	.01	-.10	-.01	.41	-							
6. Neatness of Appearance	-.08	-.06	.00	.08	.23	-						
7. Cooperation	.05	-.01	-.18	.59	.57	.22	-					
8. Watch Standing	-.29	-.02	-.32	.55	.53	.31	.64	-				
9. Knowledge of Job	-.12	.07	-.18	.37	.62	.01	.37	.42	-			
10. Discipline	.05	.04	.28	.28	.35	.37	.37	.62	.12	-		
11. Application and Initiative	.03	.09	.05	.53	.67	.21	.89	.62	.65	.45	-	
12. Judgment and Common Sense	.00	-.08	.38	.51	.66	.22	.56	.54	.77	.41	.66	-
13. Dependability	-.07	-.18	.18	.54	.74	.37	.63	.61	.56	.41	.79	.67
14. Adaptability	-.10	.12	-.22	.59	.58	.20	.60	.67	.63	.26	.62	.69
15. Leadership	.12	.00	.24	.58	.59	.18	.52	.42	.60	.47	.66	.63
16. Overall Efficiency	-.06	.08	.14	.63	.71	.34	.65	.72	.65	.40	.67	.78
17. Neatness of Work	.04	-.10	.17	.43	.69	.29	.59	.58	.47	.22	.69	.71
18. Ability to be Taught	.14	.34	-.03	.38	.66	.27	.65	.65	.69	.46	.75	.70
19. Care of Equipment	.05	.04	.10	.39	.62	.43	.58	.72	.55	.61	.64	.69
20. Ability to Troubleshoot	.35	.15	.00	.47	.66	.28	.50	.34	.53	.26	.83	.51
21. Sincerity in Job	-.22	.10	.04	.54	.67	.19	.72	.69	.59	.45	.94	.72
22. Manual Skill	.21	-.06	.12	.43	.53	.09	.46	.23	.44	.22	.48	.49
23. Overall Efficiency in Rate	.11	-.03	.24	.49	.62	.33	.67	.54	.58	.43	.85	.75
24. Gen. Classification Test	-.23	-.16	.12	-.07	-.20	.00	-.08	-.12	-.05	-.04	.00	-.20

* Tetrachoric Correlation Coefficients

TABLE VII (Continued)

Trait	13	14	15	16	17	18	19	20	21	22	23	24
13. Dependability	-											
14. Adaptability	.64	-										
15. Leadership	.64	.60	-									
16. Overall Efficiency	.81	.84	.67	-								
17. Neatness of Work	.76	.73	.51	.73	-							
18. Ability to be Taught	.70	.84	.50	.80	.76	-						
19. Care of Equipment	.64	.50	.48	.74	.62	.66	-					
20. Ability to Troubleshoot	.46	.76	.54	.78	.67	.80	.48	-				
21. Sincerity in Job	.77	.72	.63	.85	.72	.69	.70	.78	-			
22. Manual Skill	.56	.57	.39	.60	.76	.69	.49	.54	.62	-		
23. Overall Efficiency in Rate	.72	.80	.59	.77	.84	.81	.58	.77	.80	.64	-	
24. Gen. Classification Test	.00	-.09	-.19	.02	-.21	-.15	.00	-.37	-.18	-.03	-.19	-

TABLE VIII

INTERCORRELATIONS* OF RATING SCALE TRAITS, NAVY G.C.T. AND BIOGRAPHICAL INFORMATION, PAY GRADE PARTIALLED OUT STATISTICALLY (SECOND SAMPLE)
N = 286 Enginemen and Electrician's Mates

Trait	1	2	3	4	5	6	7	8	9	10
1. Age	-									
2. Length on Board	-.01	-								
3. Education	.17	-.27	-							
4. Social Adjustment	-.24	.03	.02	-						
5. Quality of Work	-.16	.18	.06	.56	-					
6. Cooperation	-.16	.19	.14	.65	.72	-				
7. Watch Standing	.03	.12	.11	.57	.80	.70	-			
8. Knowledge of Job	.00	.24	.16	.52	.77	.60	.71	-		
9. Discipline	-.24	-.02	.21	.32	.54	.48	.52	.37	-	
10. Application and Initiative	.09	.17	.11	.56	.80	.68	.72	.69	.41	-
11. Judgment and Common Sense	.06	.15	-.11	.56	.79	.57	.70	.77	.47	.63
12. Leadership	-.32	.11	.07	.60	.80	.68	.69	.84	.34	.58
13. Neatness of Work	-.07	.12	.16	.55	.77	.58	.77	.68	.32	.70
14. Ability to be Taught	.10	-.01	.20	.52	.81	.75	.76	.60	.54	.66
15. Care of Equipment	.13	.18	.04	.58	.77	.68	.76	.76	.49	.74
16. Sincerity in Job	.05	.06	.09	.60	.77	.70	.74	.70	.46	.67
17. Overall Efficiency in Rate	-.08	.16	.08	.41	.73	.53	.73	.77	.34	.58
18. Overall Efficiency	-.16	.18	.04	.55	.77	.64	.74	.79	.48	.57
19. Length in Service	.63	-.14	-.11	-.35	-.11	-.20	-.20	.07	-.64	-.08
20. Gen. Classification Test	-.28	-.04	.27	-.03	-.08	-.10	.03	-.13	.08	-.06

* Tetrachoric Correlation Coefficients

TABLE VIII (Continued)

Trait	11	12	13	14	15	16	17	18	19	20
11. Judgment and Common Sense	-									
12. Leadership	.74	-								
13. Neatness of Work	.60	.76	-							
14. Ability to be Taught	.80	.68	.68	-						
15. Care of Equipment	.68	.64	.74	.75	-					
16. Sincerity in Job	.62	.68	.71	.66	.77	-				
17. Overall Efficiency in Rate	.80	.79	.60	.72	.65	.67	-			
18. Overall Efficiency	.76	.81	.62	.69	.73	.71	.89	-		
19. Length in Service	-.05	-.16	.03	-.02	.02	.02	-.03	-.23	-	
20. Gen. Classification Test	-.12	-.02	-.03	.14	.14	.13	.04	-.02	-.49	-

TABLE IX

FACTOR LOADINGS AFTER ROTATION AND COMMUNALITIES BEFORE AND AFTER ROTATION (FIRST SAMPLE, N = 187, ACROSS PAY GRADE)

Factor	I	II	III	IV	h_u^{2**}	h_r^{2**}
Trait						
1. Pay Grade	27*	97	07	06	1.04	1.03
2. Age	22	75	-10	15	66	64
3. Length on Board	-06	36	36	11	31	31
4. Education	-11	-12	39	02	17	18
5. Social Adjustment	60	08	07	17	41	41
6. Quality of Work	56	31	56	34	63	63
7. Neatness of Appearance	34	25	40	-02	35	34
8. Cooperation	79	10	16	28	74	74
9. Watch Standing	52	44	24	55	83	82
10. Knowledge of Job	40	78	26	23	90	89
11. Discipline	51	14	45	02	49	49
12. Application and Initiative	77	14	29	26	77	77
13. Judgment and Common Sense	58	64	13	37	91	91
14. Dependability	64	59	27	26	91	90
15. Adaptability	73	45	27	22	86	85
16. Leadership	52	69	19	27	86	86
17. Overall Efficiency	69	34	32	46	90	91
18. Neatness of Work	42	27	52	59	88	87
19. Ability to be Taught	75	35	29	21	82	81
20. Care of Equipment	67	37	50	10	85	85
21. Ability to Troubleshoot	48	63	38	37	85	85
22. Sincerity in Job	69	38	29	45	91	91
23. Manual Skill	42	56	22	44	74	74
24. Overall Efficiency in Rate	48	37	60	57	94	94
25. Gen. Classification Test	-08	28	07	12	10	10

* Decimal points omitted

** h_u^2 - communalities before rotation

h_r^2 - communalities after rotation

TABLE X

FACTOR LOADINGS AFTER ROTATION AND COMMUNALITIES BEFORE AND AFTER ROTATION
(FIRST SAMPLE, N = 187, BASED ON VARIANCE WITHIN PAY GRADE

Factor	I	II	III	IV	V	h_u^{2**}	h_r^{2**}
Trait							
1. Age	13*	-21	05	07	30	16	16
2. Length on Board	-19	-11	19	07	03	09	09
3. Education	-07	32	-08	-18	15	17	17
4. Social Adjustment	57	02	32	41	-03	59	59
5. Quality of Work	30	23	32	79	16	88	89
6. Neatness of Appearance	-02	46	18	11	38	41	40
7. Cooperation	56	30	20	50	15	71	71
8. Watch Standing	35	46	19	32	36	61	60
9. Knowledge of Job	-02	37	51	55	36	82	82
10. Discipline	-08	60	30	23	-12	52	52
11. Application and Initiative	-05	42	48	27	45	69	69
12. Judgment and Common Sense	19	53	25	23	51	69	69
13. Dependability	05	41	49	33	58	86	86
14. Adaptability	32	48	19	68	13	83	84
15. Leadership	-02	50	15	54	35	69	69
16. Overall Efficiency	15	47	55	53	32	92	93
17. Neatness of Work	52	33	59	08	27	81	81
18. Ability to be Taught	20	07	55	34	40	62	62
19. Care of Equipment	28	43	28	63	18	74	77
20. Ability to Troubleshoot	-04	30	78	46	23	90	90
21. Sincerity in Job	26	46	25	35	56	80	81
22. Manual Skill	20	22	52	18	52	66	66
23. Overall Efficiency in Rate	23	54	24	62	09	80	80
24. Gen. Classification Test	-22	27	01	05	24	18	18

* Decimal points omitted

** h_r^2 - communalities after rotation

h_u^2 - communalities after rotation

TABLE XI

FACTOR LOADINGS AFTER ROTATION AND COMMUNALITIES BEFORE AND AFTER ROTATION
(FIRST SAMPLE, N = 187, PAY GRADE PARTIALLED OUT)

Factor	I	II	III	IV	V	VI	h_u^{2**}	h_r^{2**}
Trait								
1. Age	41*	33	-30	00	08	-03	37	38
2. Length on Board	00	-28	10	-07	-16	44	32	31
3. Education	-14	62	-13	23	-20	14	53	53
4. Social Adjustment	40	-06	24	28	09	44	52	50
5. Quality of Work	14	26	54	40	33	09	65	66
6. Neatness of Appearance	-01	18	22	-12	41	10	31	28
7. Cooperation	55	01	35	53	13	07	75	73
8. Watch Standing	28	05	02	41	57	57	90	89
9. Knowledge of Job	22	-03	39	46	49	-14	68	69
10. Discipline	-10	31	44	01	51	13	57	58
11. Application and Initiative	59	-03	52	64	13	-09	1.06	1.05
12. Judgment and Common Sense	45	42	50	47	06	05	84	85
13. Dependability	29	36	72	16	04	32	88	86
14. Adaptability	37	-17	74	48	01	02	91	94
15. Leadership	05	31	52	41	29	05	66	62
16. Overall Efficiency	59	12	52	36	22	30	93	89
17. Neatness of Work	26	28	34	67	11	34	84	83
18. Ability to be Taught	-02	-06	79	45	38	29	1.06	1.06
19. Care of Equipment	09	28	62	11	45	25	75	75
20. Ability to Troubleshoot	44	01	79	03	23	25	94	94
21. Sincerity in Job	26	03	45	61	39	35	91	92
22. Manual Skill	10	15	64	28	-14	30	62	63
23. Overall Efficiency in Rate	26	16	75	42	20	01	90	87
24. Gen. Classification Test	22	10	-37	-20	-03	28	32	31

* Decimal points omitted

** h_u^2 - communalities before rotation

h_r^2 - communalities after rotation

TABLE XII

FACTOR LOADINGS AFTER ROTATION AND COMMUNALITIES BEFORE AND AFTER ROTATION (SECOND SAMPLE, N = 286, ACROSS PAY GRADE)

Factor	I	II	III	h_u^{2**}	h_r^{2**}
Trait					
1. Pay Grade	99*	04	14	1.00	1.00
2. Age	95	02	10	90	91
3. Length on Board	05	04	46	22	22
4. Education	-22	19	41	25	25
5. Social Adjustment	44	62	01	58	58
6. Quality of Work	74	60	14	93	94
7. Cooperation	56	68	21	81	82
8. Watch Standing	58	70	22	87	88
9. Knowledge of Job	80	42	38	95	96
10. Discipline	32	59	27	53	53
11. Application and Initiative	59	65	10	78	78
12. Judgment and Common Sense	80	41	35	93	93
13. Leadership	78	46	33	92	92
14. Neatness of Work	73	55	15	85	85
15. Ability to be Taught	45	75	30	84	85
16. Care of Equipment	68	64	11	88	88
17. Sincerity in Job	66	64	14	86	86
18. Overall Efficiency in Rate	73	54	05	83	83
19. Overall Efficiency	72	60	06	88	88
20. Length in Service	92	-02	21	88	88
21. Gen. Classification Test	32	09	-26	18	18

* Decimal points omitted

** h_u^2 - communalities before rotation

h_r^2 - communalities after rotation

TABLE XIII

FACTOR LOADINGS AFTER ROTATION AND COMMUNALITIES BEFORE AND AFTER ROTATION
(SECOND SAMPLE, N = 286, BASED ON VARIANCE WITHIN PAY GRADE)

Factor	I	II	III	IV	V	h_u^{2**}	h_r^{2**}
Trait							
1. Age	-21*	14	70	02	-16	57	58
2. Length on Board	-02	32	03	2?	?3	14	14
3. Education	01	-01	-13	55	28	39	39
4. Social Adjustment	60	64	10	00	-01	79	78
5. Quality of Work	79	18	17	34	26	86	87
6. Cooperation	77	31	18	02	16	74	75
7. Watch Standing	46	38	16	54	06	67	67
8. Knowledge of Job	58	36	48	02	2?	73	74
9. Discipline	48	-03	18	16	31	38	39
10. Application and Initiative	72	23	17	42	02	76	77
11. Judgment and Common Sense	34	52	48	30	40	86	86
12. Leadership	57	20	31	46	07	68	68
13. Neatness of Work	60	10	15	28	42	64	65
14. Ability to be Taught	65	17	40	43	25	85	86
15. Care of Equipment	66	14	08	41	18	67	66
16. Sincerity in Job	58	26	09	29	61	84	85
17. Overall Efficiency in Rate	78	18	16	15	38	84	85
18. Overall Efficiency	62	41	20	50	16	86	87
19. Length in Service	-07	-56	28	-07	21	44	45
20. Gen. Classification Test	10	21	-07	-24	39	27	27

* Decimal points omitted

** h_u^2 - communalities before rotation

h_r^2 - communalities after rotation

TABLE XIV

FACTOR LOADINGS AFTER ROTATION AND COMMUNALITIES BEFORE AND AFTER ROTATION
(SECOND SAMPLE, N = 286, PAY GRADE PARTIALLED OUT)

Factor	I	II	III	IV	V	VI	VII	h_u^{2**}	h_r^{2**}
Trait									
1. Age	-07*	34	-61	-24	20	-26	-30	73	74
2. Length on Board	23	-03	38	-23	-14	-15	32	39	40
3. Education	05	31	33	-22	-13	02	17	30	30
4. Social Adjustment	35	16	16	40	10	31	34	50	57
5. Quality of Work	73	09	29	24	44	12	11	91	90
6. Cooperation	53	16	01	41	31	07	41	75	75
7. Watch Standing	58	09	13	23	54	31	14	83	82
8. Knowledge of Job	72	-01	42	31	31	13	09	89	90
9. Discipline	18	55	06	04	50	27	05	66	66
10. Application and Initiative	62	-01	04	34	20	26	36	74	74
11. Judgment and Common Sense	33	-01	39	42	49	36	27	88	88
12. Leadership	45	24	47	18	62	02	23	95	95
13. Neatness of Work	57	-01	10	46	42	14	12	76	76
14. Ability to be Taught	88	03	-01	01	37	03	15	94	93
15. Care of Equipment	67	-05	01	23	54	01	31	90	89
16. Sincerity in Job	55	00	00	34	49	36	09	81	80
17. Overall Efficiency in Rate	52	09	47	48	10	36	28	93	94
18. Overall Efficiency	55	21	45	42	26	27	14	89	89
19. Length in Service	-17	64	-51	-2[illegible]	-40	08	23	98	97
20. Gen. Classification Test	-09	53	-03	-37	-02	27	23	55	55

* Decimal points omitted

** h_u^2 - communalities before rotation

h_r^2 - communalities after rotation

APPENDIX B

PERFORMANCE RATING SCALE

FOR ENLISTED NAVAL PERSONNEL

(OFFICE OF NAVAL RESEARCH PROJECT N8 onr 70001)

U.S.S._____

DIVISION:_____

NAME OF RATER:_____

RANK OR RATE:_____

DATE:_____

Form RF 108

Psychological Research Center

Los Angeles, California

KNOWLEDGE OF THE JOB

How well does each man know the technical duties of his rate? Does he know all the fine points or does he just know enough to get by? How often does someone have to help him get the job done? Does he know more or less than most men of the same rate?

	E.J.Brownell EM 2/c	J.A.Carson EM 3/c	L.L.Furgeson EM 1/c	T.J.Harrison EM 1/c	F.G.Moore EMFA	F.L.Parker EM 2/c	W.N.Thompson EM 3/c	S.E.Turner EMFA
Has as much or more job knowledge as the very best of his rate. Knows all the fine points of his job. Never needs help on the job.			√					
Practically never needs help on the job. Has very little to learn about his job except for a few of the most technical aspects.	√			√				
Has adequate knowledge to turn in a satisfactory job. Sometimes needs help on the job, since he still has a few things to learn.						√	√	
Has much to learn about his job. Needs supervision very frequently since he lacks much of the fundamental knowledge necessary to do the job alone.		√			√			√

INSTRUCTIONS TO RATERS

You will find the names of the men in your gang or division listed in this booklet. They are to be rated by you on several qualities which you will recognize as important for success in the Navy.

The qualities are described at the top of each page. Along the side of each page various degrees of the quality in question are indicated. Some men in your group will be pretty well described by one particular group of statements, while others may seem to fall between two groups of statements (that is, two groups of statements partially describe them.) Your task is to indicate with a check mark the position of each man in your group with respect to the trait in question.

On the following page a sample rating sheet is shown to indicate how a completed rating might look. Here each man has been compared with the others and rated on the trait of **Job Knowledge.** Notice that the check marks have been well spread out and that no two people received the same rating. These are two characteristics of well-performed ratings.

Observing the following procedures will help make your ratings most valuable:

1) Consider each man in comparison to the others in his group as you rate each trait. There will be a few outstanding men in each gang, and a few who are not so good, while most will fall somewhere in between.

2) Rate the best and the poorest man in each trait first. With these limits established, rate the remainder of the men in relation to them.

3) If possible, avoid giving tie ratings. Before you rate two men exactly equal, consider carefully whether there is not some small difference between them.

4) Remember that most men's ratings vary from trait to trait. Because a man is high (or low) in one trait does not necessarily mean he will be high (or low) on the others. Each trait should be considered separately. For example, it is improbable that the same man or men will be highest (or lowest) in **every** trait in the scale.

5) **This is important.** Do not rate one man higher than another simply because he is in a higher pay grade. Experience and ability generally run together, but they are not always perfectly related. These ratings are to be used for research purposes only and will not be a factor in anyone's advancement. We sincerely request your frank and honest impressions.

COOPERATION

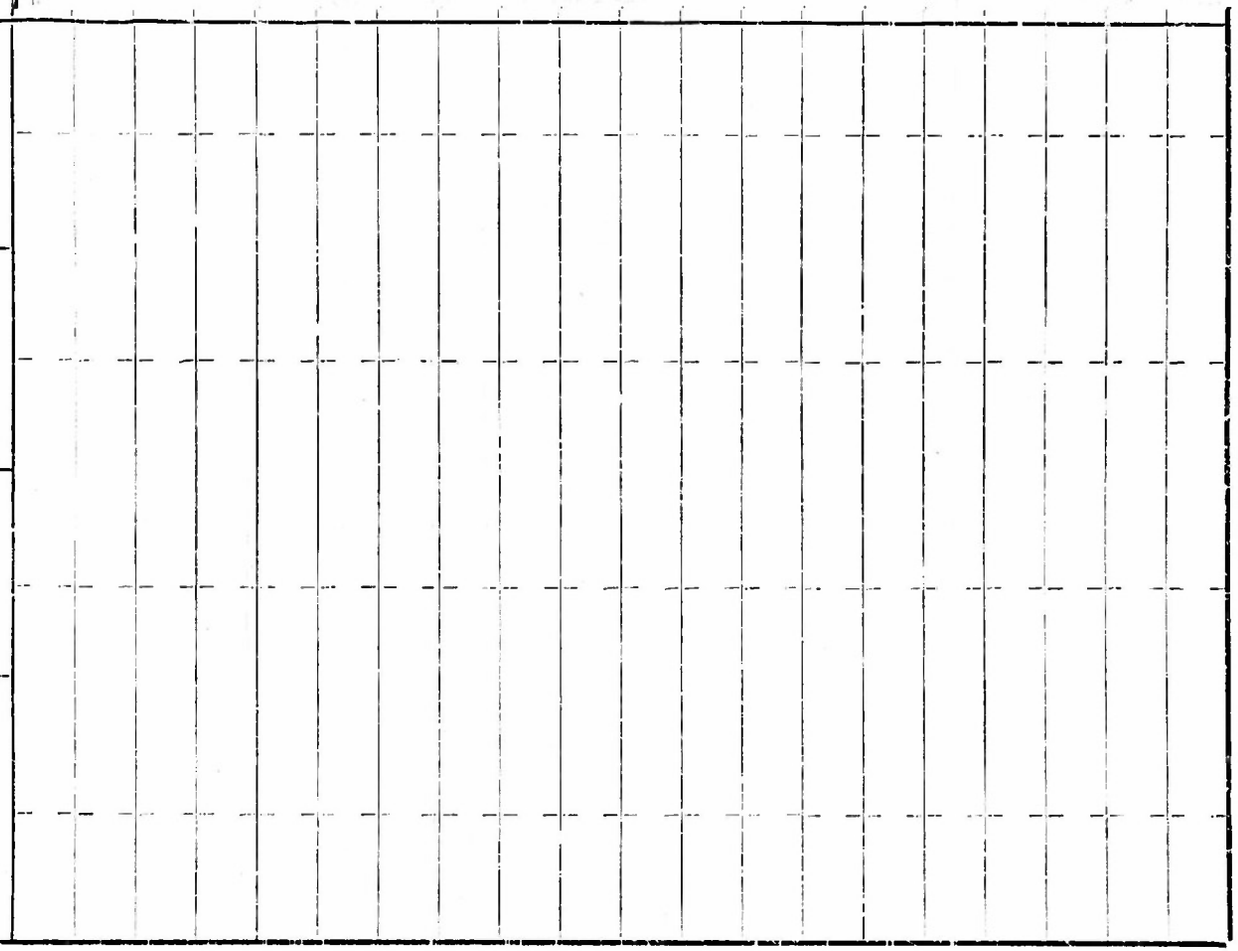
How well does each man work with others? Is he all for himself or does he willingly pitch in and help other members of his gang? Does mutual benefit mean more to him than his own personal ambitions and pleasures?

About the most cooperative man on board. Goes out of his way to help others. Accepts orders willingly and cheerfully. Never passes the buck.

Never causes any trouble over working with others. Accepts orders willingly and promptly carries them out. Very willing to help others.

Reasonably cooperative in most respects. Accepts orders without griping. Does not go out of his way to help others, but is not strictly for himself either.

Never does anything for anybody unless ordered. Often passes the buck. Tries to avoid the job. All out for himself.



KNOWLEDGE OF THE JOB How well does each mar know the technical duties of his rate? Does he know all the fina points or does he just know enough to get by? How often doas someone have to help him get the job clone? Does he know more or less than most men of the same rate?	
Has as much or more job knowledge as the very best of his rate. Knows all the fine points of his job. Never needs help on the job.	
Practically never needs help on the job. Has very little to learn about his job except for a few of the most technical aspects.	
Has adequate knowledge to turn in a satisfactory job. Sometimes needs help on the job, since he still has a few things to learn.	
Has much to learn about his job. Needs supervision very frequently since he lacks much of the fundamental knowledge necessary to do the job alone.	

DISCIPLINE How often does each man get himself into trouble? (Consider conduct both ashore and afloat). Has he ever had formal disciplinary action? Has he ever been restricted? Does he have to have verbal warnings about his conduct or the way in which he responds to orders?	
Never gets into trouble either ashore or afloat. Obeys all regulations and responds to orders without question. Never needs to be spoken to.	
Hardly ever gets into trouble. May break a rule once in a great while, but is quickly straightened out if given a warning.	
Conduct is usually okay, but gets into trouble once in a while. Easily straightened out most of the time.	
Quite often gets into trouble. Often ignors warnings and has to have formal disciplinary action. Conduct frequently is a real problem.	

APPLICATION AND INITIATIVE

Consider how industrious each man is. Does he usually keep busy or does he often run out of work? Does he tackle a job on his own, or does he always have to be told to do a job? When on the job does he really put out or does he exert a minimum amount of effort?

Always busy doing something. Works extremely hard. Looks for extra work without being told. A regular workhorse.	
Keeps busy without being told. Works very hard and frequently does more than is required of him.	
Does his share of the work but seldom more than he is told to do. A good average worker.	
Frequently tries to get out of work. Needs to be pushed to get the job done. Fluffs off whenever possible.	

JUDGMENT AND COMMON SENSE

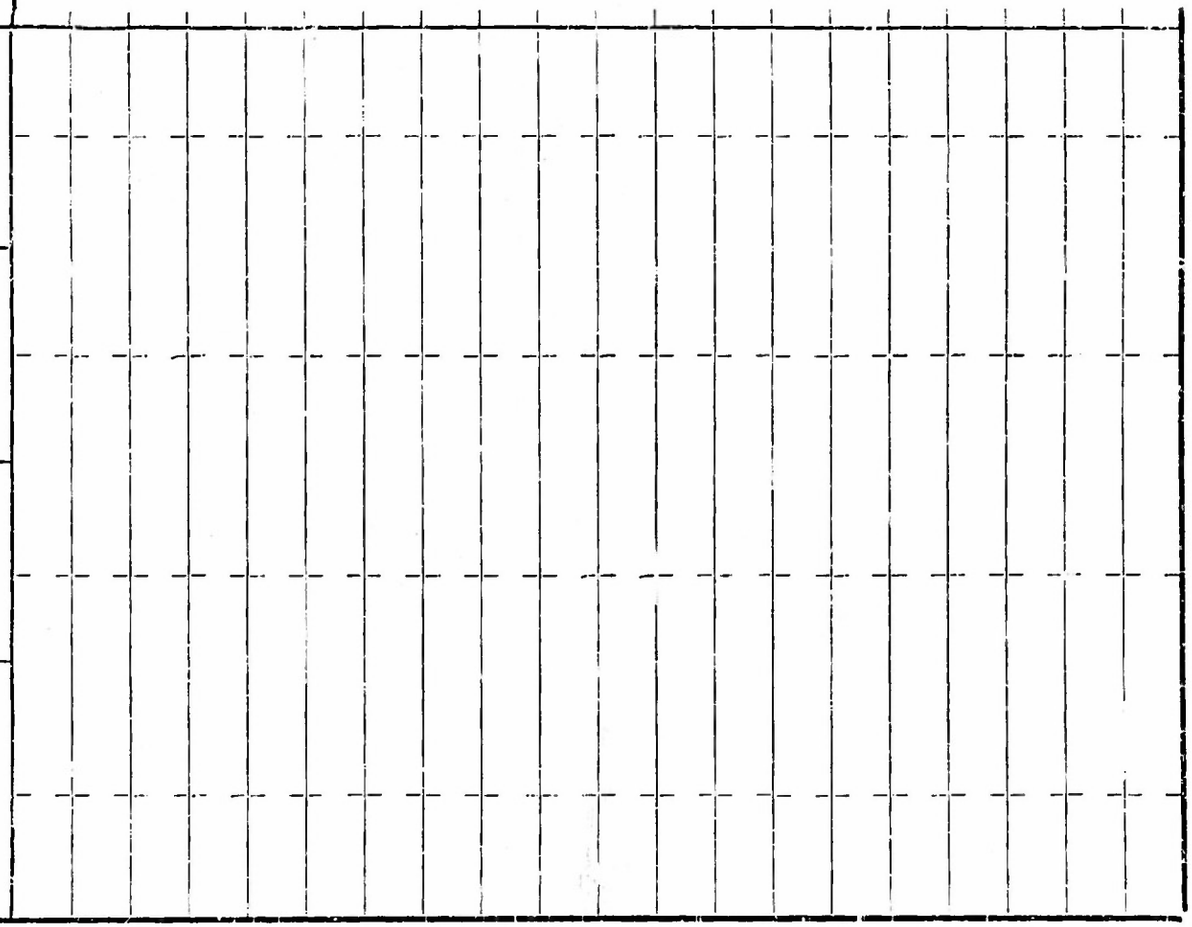
How well can you depend on each man to make correct decisions? Would you trust his judgment in an emergency? When placed on his own can he figure things out for himself and make proper decisions, or do you frequently have to check to see that he has not fouled things up?

Can rely completely on his ability to make proper decisions. Can be depended upon to handle almost any emergency. Never fails to display excellent judgment.

Has very good judgment and seldom runs into a situation which he cannot handle. Needs advice only occasionally on what is the proper decision.

Usually uses common sense, but does not have a particularly mature judgment. Needs advice fairly frequently in making decisions.

Sometimes makes unsound judgments or decisions. Often does not use common sense. Cannot rely very highly on his ability to decide correctly.



LEADERSHIP Consider whether each man is primarily a leader or primarily a follower in his gang. Does he tend to be forceful? Do others respect him? (Is the response to his orders rapid and cheerful?) Is he looked up to by men of his same rate?	
Among the most respected and admired men in the boat. A forceful leader who gets things done. Men willingly follow his advice and orders.	
Tends to be a leader. Often directs the work of his gang, and is usually well accepted. Is sufficiently respected and admired to get response from the men.	
Occasionally displays some leadership, but usually just a good follower. Well liked by the gang, but not particularly looked up to.	
Always a follower. Never has any suggestions, and is not looked to for advice. Shows no ambition to lead others.	

ABILITY TO TROUBLESHOOT

How quickly and efficiently can each man locate the difficulty when a piece of equipment breaks down? Does he seem to know just where to look for the trouble or do breakdowns and similar problems throw him for a loss?

No problem or breakdown is too tough for this man. Always knows exactly where to go to locate the difficulty, and is very quick in making needed repairs.

A very good troubleshooter. Few problems are too tough for him, given sufficient time to smell out the difficulty. Almost always knows where to look.

Needs quite a bit of help in locating trouble spots. Is a reasonably good repair man, but is not outstanding in this ability.

Thrown for a loss when it comes to solving problems or fixing breakdowns. Never seems to have any idea where to look for the source of trouble.

SINCERITY IN DOING A GOOD JOB

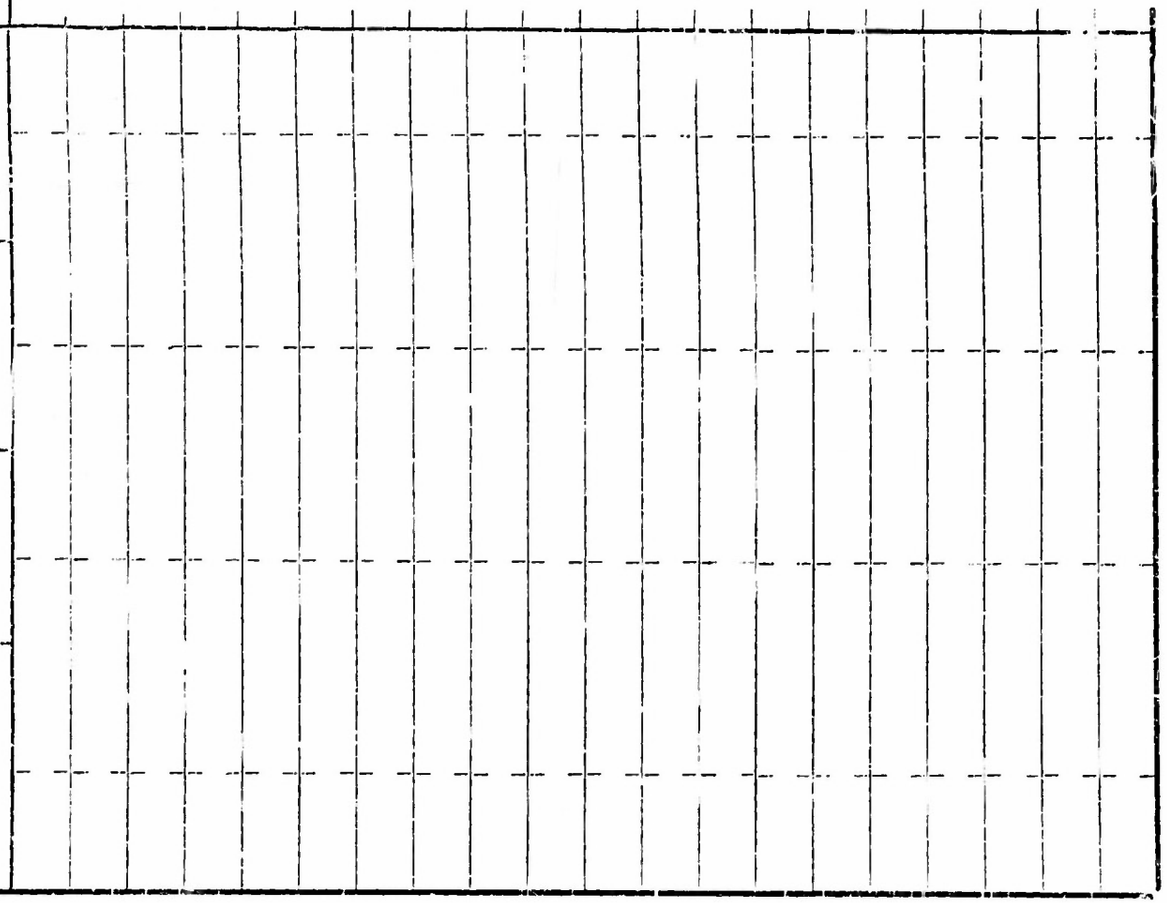
How sincere is each man in turning out the best job of which he is capable? Does he stick to the job until it is done in the best possible manner, or is he content to get by with something less than the best? Does he take pride in his work or is it just something that has to be done?

Always stays with the job until he has done the best possible job of which he is capable. Takes great pride in seeing a good job done.

Very seldom settles for anything but the best possible job. Takes considerable pride in his work. Practically never turns out a second-rate job.

Generally does a satisfactory job, but sometimes does not follow through. Occasionally does not do the job of which he is capable.

Often turns out a sloppy job. Hurries through the work with little regard for the quality of the job. Takes little pride in doing a good job.



INSTRUCTIONS:

1) If you were transferred to another ship from the one you are now on and were to be the division officer or senior petty officer, you would be interested in both the petty officer qualities (leadership, cooperation, initiative, sincerity, etc.) of your men and in their ability in rating (technical skill).

Considering both petty officer qualities and ability in rating, which one of these men would you most want to take with you if you were transferred to another ship? Which man would be of greatest value to you?

Indicate your first choice by placing number 1 opposite the man's name in the column headed "Order of Choice." Place a 2 opposite your second choice, a 3 opposite your third choice and so on until you have assigned a number to each man in the gang.

2) In the space headed "Remarks," make any comment that you feel should be considered in evaluating the ratings you have assigned to the men in your gang. For example,

"Only been aboard 60 days"
"Stands only look-out watches"
"Will never make a submarine sailor"
"Finest man I have ever known"

REMARKS	ORDER OF CHOICE	NAME AND RATE