MEDICAL RESEARCH LABORATORY REPORT NO. 73-a

LIBRARY
MEDICAL RESEARCH LABORATORY
ACCESSION NO.

AN EVALUATION OF THE PERSONAL INVENTORY AND CERTAIN OTHER MEASURES IN THE PREDICTION OF SUBMARINE OFFICERS' EVALUATIONS OF ENLISTED MEN

OSRD Report No. 5557

September 7, 1945

Applied Psychology Panel, NDRC

APPROVED FOR PUBLIC
RELEASE - DISTRIBUTION
UNLIMITED

Applied Psychology Panel, NDRC

Project N-106: RESEARCH AND DEVELOPMENT
OF THE NAVY'S APTITUDE TESTING PROGRAM

Report No. 22 September 7, 1945

AN EVALUATION OF THE PERSONAL INVENTORY AND CERTAIN OTHER MEASURES IN THE PREDICTION OF SUBMARINE OFFICERS' EVALUATIONS OF ENLISTED MEN

Prepared by

G. A. Satter

Office of Scientific Research and Development
Contract OEMsr-705

OSRD Report No. 5557 Copy No. 78 of 125 copies

Research and Statistical Laboratory
College Entrance Examination Board
Princeton, New Jersey

Harold Gulliksen, Contractor's Technical Representative
and Project Director

ACKNOWLEDGMENTS

The data which form the basis of this Report were collected through the joint efforts of Project N-113 and the officers attached to the Medical Research Laboratory at the New London Submarine Base. Special acknowledgment is given to Captain C. W. Shilling for his role in initiating and expediting the activities involved in collecting these data.

Thanks are also given to Lieutenant (jg) Neil R. Bartlett and to Dr. C. H. Graham for their responsibilities in collecting the data and for their advice and cooperation in carrying out the present analyses.

ABSTRACT

During 1943 and the early part of 1944, men processed at the New London Submarine Base were given a battery of tests which included a Two-Hand Coordination Test, the Otis Self-Administering Tests of Mental Ability, the Personal Inventory, the New London-NDRC Confidential Questionnaire, and the NRC Neurotic Inventory. These test scores, along with school grades and a set of trait-ratings made by submarine officers, were later made available to Project N-106 for study and analysis. This analysis involved the correlation of test scores and school grades with a selection of trait-ratings in order to determine the predictive value of these measures.

As a result of these analyses the following conclusions were reached:

1. Scores on the Personal Inventory and the other selection measures were found to be unrelated to quality of performance aboard submarines as it was rated by officers. In a similar manner, an analysis of psychiatrists' evaluations of men judged to be "emotionally unstable" suggests that, for this sample at least, these evaluations are no more closely related to officers' ratings than those made by the paper and pencil tests.

   Evidence gleaned from earlier investigations would suggest that the present findings are by no means typical -- the PI, as well as the other measures of the selection battery, has been shown to be capable of making reliable discriminations among certain types of groups. The lack of correlation reported in the present study can be attributed to the fact that (a) the sample dealt with is select with respect to mental, physical, and emotional characteristics; (b) the criterion is undoubtedly quite unreliable; and (c) the criterion is not entirely appropriate.

2. An analysis of psychiatrists' evaluations of men judged to be "emotionally unstable" suggests that, for this sample at least, these evaluations are no more valid than those made by the paper and pencil tests.

3. Grades made in the submarine training courses can, to a certain extent be used in predicting ratings made by submarine officers. Likewise, certain of the advanced school grades --those received after Diesel, Battery and Gyro, Ordnance, Sound, and Radio Training -- also offer possibilities as predictors.

4. Since grades offer the most by way of possibilities for improved prediction, it is suggested that further work be directed toward the improvement of both the school grade and the criterion.

AN EVALUATION OF THE PERSONAL INVENTORY AND CERTAIN OTHER MEASURES IN THE PREDICTION OF SUBMARINE OFFICERS' EVALUATIONS OF ENLISTED MEN

I. INTRODUCTION

The Personal Inventory and other psychological tests employed at the U. S. Submarine Base at New London, Connecticut, have been evaluated by members of the staff of NDRC Project N-113 working cooperatively with the officers of the Medical Research Laboratory at the Base.[1] In making these studies various criteria were used: psychiatric evaluations, officers' ratings of experienced submarinemen, and ratings of escape tank performance. The present analysis extends these evaluations to a number of other tests and to the grades assigned submarinemen at the termination of their basic and advanced training; in this instance, ratings made by submarine officers are used as criteria.

II. THE DATA

The data which form the basis of this Report were taken from the card files of the Medical Research Laboratory at the New London Submarine Base where the men had been processed. When the work of Project N-113 was terminated, these data were, at the request of the Applied Psychology Panel, turned over to Project N-106 for analysis and study.

These record cards provide some twenty-eight items of information (test scores, grades, etc.) which include:

1. Age -- recorded as of last birthday at the time of testing.

2. Grade in Basic Training -- recorded in terms of the 0-4 Navy scale.

3. Class standing in Basic Training -- recorded in terms of percentage position as computed by means of the formula $\frac{(\text{rank} - .5)\ 100}{N}$.

4-20. Grade in Advanced Training -- recorded in terms of the 0-4 Navy scale for the following advanced schools:

4. Diesel	10. Sound	16. Quartermaster
6. Battery and Gyro	12. Radio	18. Pharmacist's Mate
8. Ordnance	14. Radar	20. Lookout

1. NDRC Project 44, Div.7, A Report of Research on Selection Tests at the U.S.Submarine Base, New London. OSRD Report No. 1770. Providence, R.I., Brown University, 1943. Pp.25. NDRC Project N-113, Final Report in Summary of Work on the Personal Inventory and Other Tests. OSRD Report No.3963. Providence, R.I., Brown University, 1944. Pp.13

5-21. Class standing in Advanced Training -- recorded for each of the nine schools in terms of percentage position as computed by means of the formula $\frac{(\text{rank} - .5)100}{N}$.

22. Ratings on Escape Tank Performance -- recorded in terms of a 1 to 5 rating.

23. Score on the Two-Hand Coordination Test. Scores represent the total amount of time "spent in contact" on a motor pursuit task.[2]

24. Score (IQ) on the Otis Self-Administering Tests of Mental Ability. Higher Examination: Form B.[3]

25. Score on the Personal Inventory, Format B. This is a group test of 145 items which Project N-113 developed for use in the preliminary screening of "psychiatrically undesirable men".[4]

26. Score on the New London-NDRC Confidential Questionnaire. This questionnaire, which samples various aspects of the individual's background, was developed at New London and validated on a submarine school group by using psychiatric evaluations and tank performance as criteria.[5]

27. Score on the NRC Neurotic Inventory, Form R-2. This inventory consists of forty-seven items sampling "likes and dislikes" and "wishes and worries".[6]

28. Ratings of proficiency made by submarine officers. By directive, ratings were made on a seven-trait graphic scale and were returned to the Medical Research Laboratory where they were entered on the record card. (See Illustration 1.)[7] These ratings were collected with the end in view of later using them in the evaluation of the selection measures.

These items of information, in more or less complete form, were made available for some 1600 men representing 40 submarines of the Atlantic Fleet and 58 of the Pacific. These men (a) had been processed at the Base during 1943 and the early part of 1944, (b) were sent to sea as members of a submarine crew, and (c) were rated on the seven-trait graphic scale after having completed at least one war patrol. Thus, for this group of 1600 men the New

---

2. NDRC Project 44, Div.7, A Report on Research on Selection Tests at the U.S.Submarine Base, New London. OSRD Report No. 1770. Providence, R.I., Brown University, 1943. P. 5.
3. Otis, A.S., Otis Self-Administering Tests of Mental Ability. Higher Examination, Form B. Yonkers-on-Hudson, N.Y., World Book Co., 1922.
4. NDRC Project N-113, op. cit., p. i.
5. NDRC Project 44, Div. 7, op. cit., p. 5
6. Ibid, p.5
7. See Section III, "The Criterion," for a description and evaluation of the scale.

London records provided enough data to permit the evaluation of test scores and grades as predictors of the ratings later made by submarine officers.

III. PROCEDURE

The procedure which was followed in evaluating the predictive value of the measures described above is, in general, as follows:

(1) The ratings made by the submarine officers of the 98 vessels were studied for the purpose of selecting those vessels whose ratings were best suited for use as criteria.

(a) All the ratings from each of the vessels were tabulated.

(b) On the basis of these tabulations, those vessels whose ratings lacked "spread" (where a large proportion of the ratings fell in one category of the scale) were discarded from the sample.

(c) Intercorrelations among the trait ratings from those vessels accepted under "b" above were computed and factor analyzed.

(d) The results of the factor analysis were used as the basis for selecting certain of the trait ratings for use as criteria in studying the validity of the selection and achievement measures.

(2) Test scores and school grades were correlated with the ratings on the selected traits. In this analysis the Atlantic and Pacific fleets were treated independently.

The findings which resulted from carrying out these steps are discussed in the next two sections. The study of the criterion measure and the selection of ratings is described in Section IV below, and statistical evaluations of the selection tests and school grades are presented in Section V.

IV. THE CRITERION

In evaluating the measures used as criteria in this study, it is well to bear in mind the form which these evaluations took and the procedures followed in collecting them. The ratings, by directive, were made and reported on a standard form (see Illustration 1). This form required the submarine officer to evaluate his men on seven traits presumed to be related to performance on board the submarine. These traits were: (1) Adaptability, (2) Alertness, (3) Courage, (4) Sociability, (5) Leadership, (6) Dependability, and (7) Adequacy of Overall Performance. The officer made his evaluations on a three-point graphic scale, the

Return to: **Medical Research Laboratory**
Submarine Base
New London, Connecticut

This report covers ____ weeks of patrol, and was completed by the {Commanding Officer / Executive Officer} (Check which). Make report out in ink.

_____ _____ _____
Name Service No. Rate

_____ _____, 1944.
Ship Date

1. Adaptability To Discipline: (Willingness to obey orders; cooperativeness.)	Insubordinate; disciplinary problem.	Obeys without comment; does no more nor better than he must.	Follows orders with despatch and interest.
2. Alertness: (Industry, Initiative, Self-Improvement.)	Highly interested; unusual effort to improve self.	Normal industry and interest.	Not very alert; tends to be satisfied with his lot.
3. Courage: (Emotional Stability; Reaction to Stress.)	Unruffled by any danger; able to perform well in all situations.	Excited by danger; but performs satisfactorily.	Tends to go to pieces under stress.
4. Sociability: (Ability to get along with other men.)	Extremely well liked; Good mixer; makes friends easily.	Liked quite well by most.	Few friends; generally avoided.
5. Leadership: (Degree to which respected and followed.)	Ineffective in groups; always follows lead of others.	Occasionally a leader; fairly well considered.	Lead followed by all; natural leader.
6. Dependability: (Responsibility and Reliability.)	Can be relied upon to perform well under any circumstances and at any time.	Reliable under most circumstances; rarely disappoints.	Unreliable; cannot be depended upon.
7. Adequacy of Overall Performance:	Inferior man; would like to replace.	Satisfactory; desire to retain.	Extremely valuable; special desire to retain.

_____
Your Initials Here

ADDITIONAL COMMENT _____

_____

_____

_____

Illustration 1: Rating Scale Form Used in Collecting the data for the Criterion Measure.

points of which were defined by adjectives and short descriptive phrases. On completing the ratings, the officer initialed the rating form and returned it to the Medical Research Laboratory at New London. Sixteen hundred of these rating forms were turned over to Project N-106 for study and analysis.

Under the procedures described above, it should be noted that it was not possible to train the officers in rating procedures or to review and follow-up their evaluations periodically. This would lead one to anticipate (in terms of the experiences of previous investigators) that most of the errors which are commonly made by untrained raters would be present in varying amounts in our sample of ratings and that these errors would reduce the usefulness of the ratings as criteria. Steps were thus taken to discard those ratings which were obviously worthless as measures of performance. The procedure adopted was as follows: The ratings received were tabulated for each submarine and the number of ratings falling in each category of the scale was determined. These tabulations revealed that there was very little spread (most of the ratings fell in the middle category of the scale) in the ratings made by some submarine officers. Quite arbitrarily, but in the interest of improving the criterion, ratings from 9 Atlantic and 6 Pacific vessels were discarded. The standard was based on the percentage of ratings in the extreme categories (see Figure 1). If less than 26 percent of the ratings of the men on a vessel fell in the extreme categories, the ratings for all the men from such a vessel were discarded. In all, the ratings for 150 men from the Atlantic fleet and 67 from the Pacific were rejected for this reason. Further, some of the vessels supplied so few ratings that it was felt advisable to discard these vessels from the sample on the ground that valid ratings are difficult to make when the frame of reference in which the ratings are made is defined by a small number of men (see Figure 1). Ratings from two of the vessels (10 men) from the Atlantic fleet and 11 (42 men) from the Pacific were discarded on this basis. Thus, by applying these two criteria in selection the total number of ratings available for use as criteria was reduced from 1600 to 1400.

The effect of this selection on the statistical characteristics of the ratings is illustrated in Tables 1 and 2. In general, the effect of selection was to increase the spread of the ratings (compare traits of Tables 1 and 2) and, as one might anticipate, the intercorrelations between the individual trait ratings.

The magnitude of these intercorrelations raised the question as to whether or not each of the rating scale traits should be used as an independent criterion in the validity analysis. Obviously, if the correlation between traits is high, there is little to be gained by treating each of these traits as independent. To rationalize a selection of traits to be used, a factor analysis of the rating scale traits was made. Although two factors were extracted from the matrix (see Table 3), one of the factors contributes very little to the variance of any of the traits and should, perhaps, not have been extracted. Presumably, a great deal of "halo" characterizes the individual trait ratings and one could choose almost any

TABLE 1

Intercorrelations* of Rating Scale Traits
(Accepted Ratings)

Trait	2	3	4	5	6	7	Sum	M	σ
(1) Adaptability	.523	.293	.388	.393	.493	.480	.720	2.63	.512
(2) Alertness		.329	.427	.487	.593	.624	.788	2.34	.622
(3) Courage			.364	.374	.489	.416	.599	2.38	.526
(4) Sociability				.412	.495	.512	.670	2.32	.53[illegible]
(5) Leadership					.500	.457	.714	1.94	.51[illegible]
(6) Dependability						.671	.819	2.33	.545
(7) Overall rating							.811	2.22	.543
Sum of trait ratings								16.15	2.782

TABLE 2

Intercorrelations* of Rating Scale Traits
(Rejected Ratings)

Trait	2	3	4	5	6	7	Sum	M	σ
(1) Adaptability	.474	.168	.211	.289	.371	.337	.693	2.32	.476
(2) Alertness		.231	.289	.617	.470	.627	.782	2.12	.502
(3) Courage			.273	.236	.194	.236	.463	2.08	.308
(4) Sociability				.239	.289	.145	.518	2.11	.324
(5) Leadership					.441	.546	.593	1.94	.306
(6) Dependability						.469	.642	2.09	.330
(7) Overall rating							.678	2.03	.302
Sum of trait ratings								14.69	1.639

* These values are coefficients of contingency.

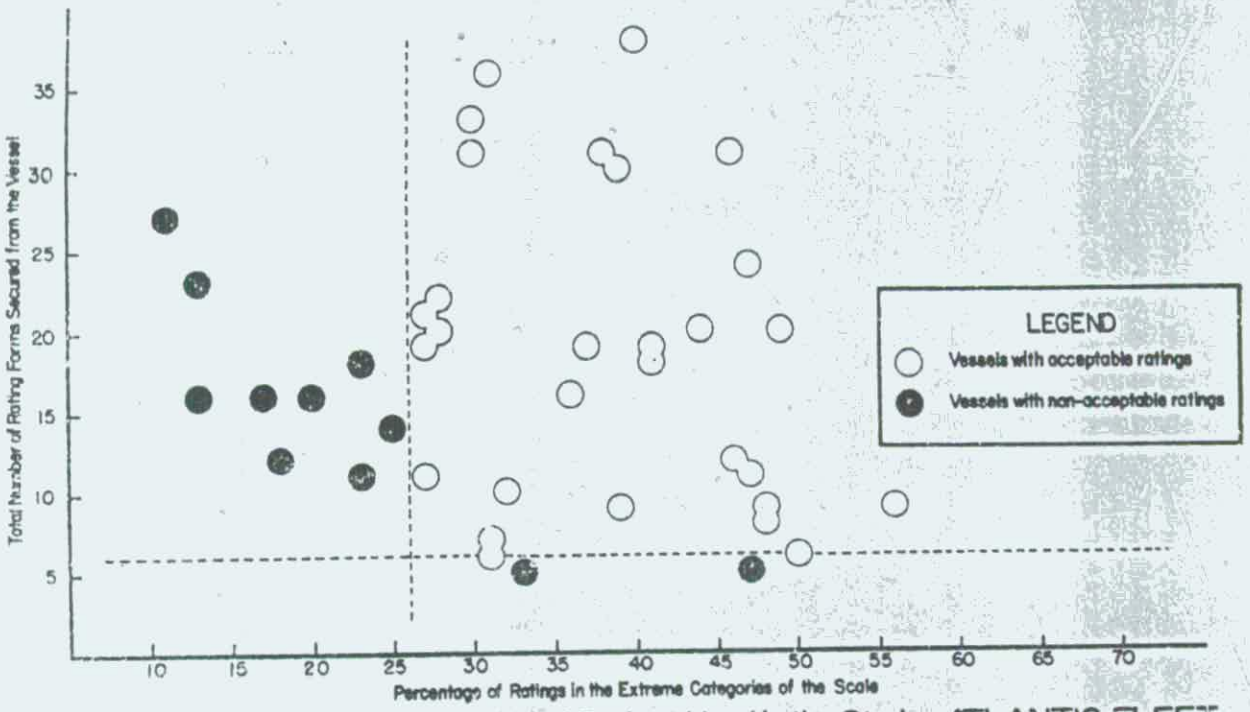

Fig. 1a: Plot Showing the Selection of Ratings Used in the Study—ATLANTIC FLEET

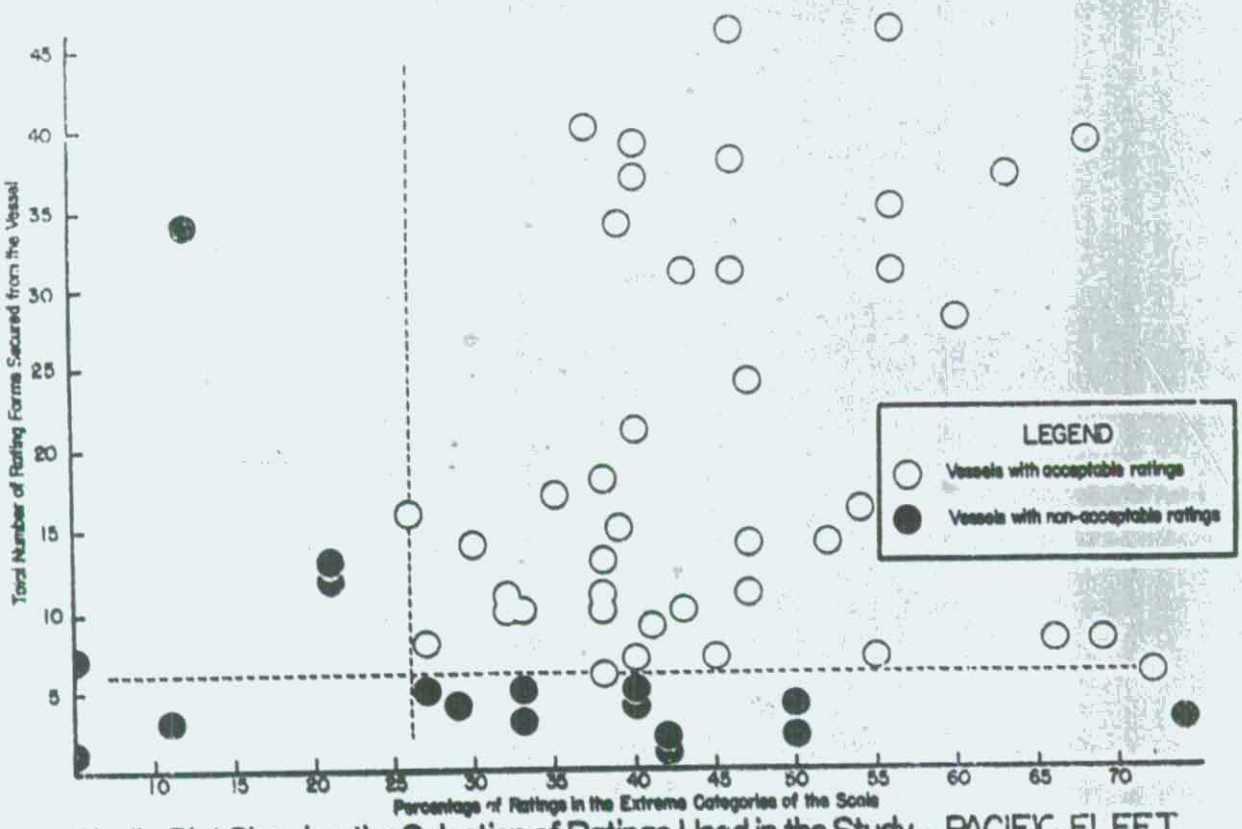

Fig. 1b: Plot Showing the Selection of Ratings Used in the Study—PACIFIC FLEET

one of the traits to represent whatever abilities, performances, etc. are described in these ratings. Table 3 also shows that only about half of the variance in the trait ratings (see h^2 column) is accounted for by the operation of the two factors which were extracted; it would seem that either the ratings are very unreliable or that one must call upon specific factors to account for the total variance in the ratings. In terms of what is known about ratings in general the latter explanation seems the less tenable.

TABLE 3

Factor Loadings and Communalities
of the Rating Scale Traits

Trait	k_1	k_2	h^2
(1) Adaptability	.639	-.221	.457
(2) Alertness	.745	-.238	.612
(3) Courage	.569	.228	.376
(4) Sociability	.642	.097	.422
(5) Leadership	.645	-.100	.426
(6) Dependability	.808	.112	.665
(7) Overall Rating	.791	.090	.633

Figure 2 shows that it is impossible to define two distinct clusters of traits to represent the two factors extracted -- one would hope for two clusters of points falling at right angles to each other -- and that actually there is very little justification for choosing one of the traits in preference to another, as a validation criterion. As a precaution, three traits (1, 3, and 7) as well as the sum of all seven trait ratings, were chosen to represent the abilities and performances measured by the scale. This procedure made it possible to eliminate approximately half of the work involved in correlating each of the traits with the selection tests and with the school grades. An exception was made in the case of the Personal Inventory; scores on this test were correlated with each of the seven trait ratings and also with the sum of the seven traits.

Thus, the analyses described in this Section resulted in a selection of ratings with an improved spread, and a set of three traits which can be taken to represent the entire scale of seven.

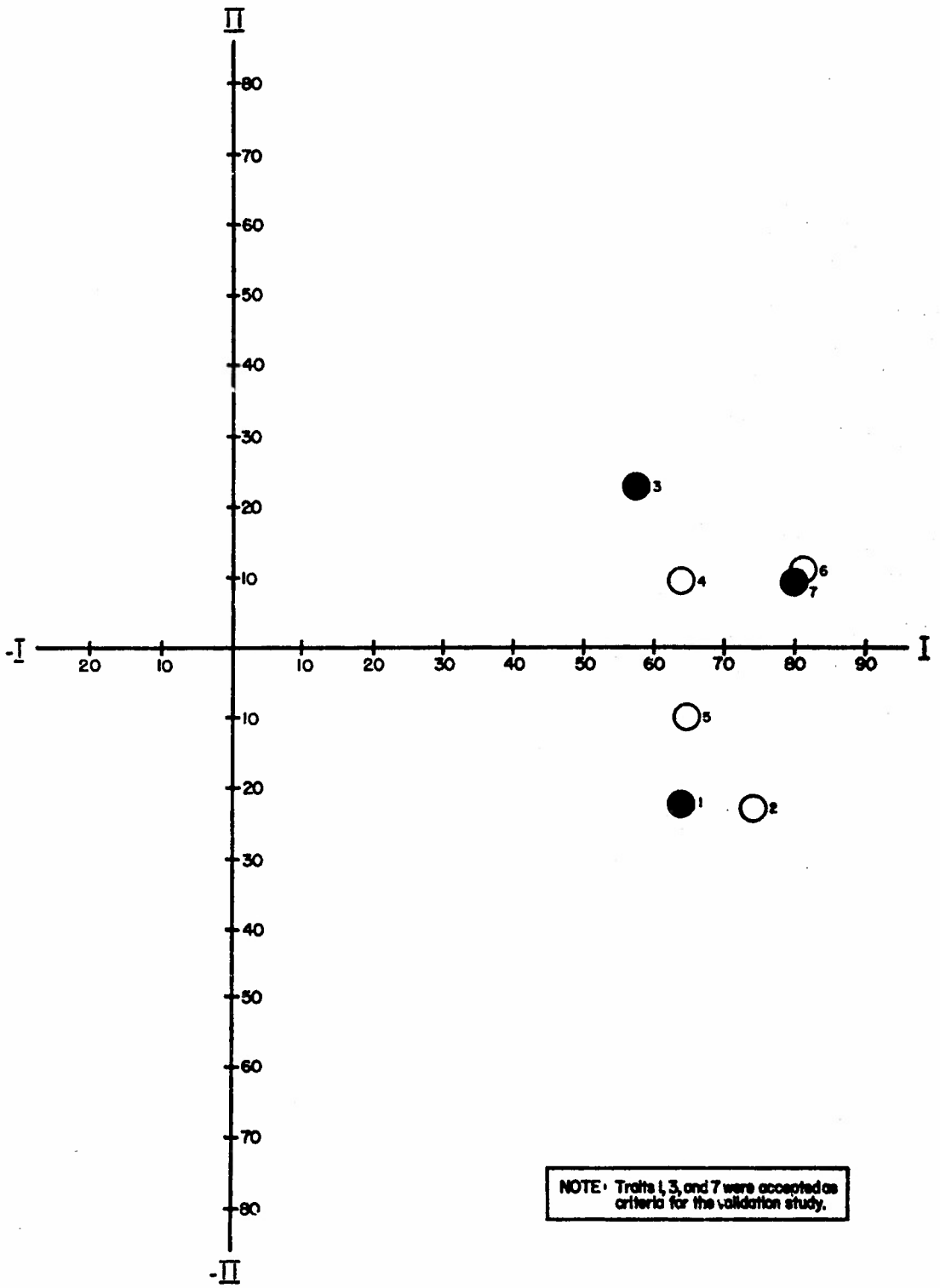

Fig 2: Plot Showing the Differences among the Seven Rating Scale Traits with Respect to their Factor Loadings

V. RESULTS

A. The Validity of the Personal Inventory

Scores on the Personal Inventory were correlated with the trait ratings transformed into numerical values. The resulting coefficients are reported in Tables 4 and 5.[8] It is difficult to arrive at any conclusion other than that scores on the Personal Inventory are unrelated to performance as it is described by officers' ratings. It is conceivable that, since the men were in part selected on the basis of PI scores, the low validity coefficients are the result of a curtailed range of adjustment; this is in part borne out by the data at hand. For the Atlantic Sample the PI distribution had a mean of 7.78 and a sigma of 4.04; for the Pacific, mean and sigma were 6.94 and 3.85, respectively. Both of these statistics are somewhat lower than those which NDRC Project N-113 has reported[9] for other Navy groups but hardly enough so as to account for the zero validities obtained.

It might also be argued, and perhaps more reasonably, that the criterion is inappropriate in the sense that the PI was not designed to make the type of discrimination which the present criteria demand -"mental health", "personal adjustment," "neurotic tendency", or whatever facet of personality structure which is tapped by the Personal Inventory --is only indirectly, and not very accurately, reflected in the trait evaluations made on the rating scale. Of course, this type of argument can be used to rationalize any table of low validity coefficients but in this particular case it would seem more than a rationalization. The PI was validated originally on a psychiatric criterion; and in successive trials, when similar criteria were employed, the instrument stood up reasonably well.[10] When many investigators frequently have had difficulty in making personality measures stand up from one sample to another, it would seem unjust to expect the Personal Inventory (or for that matter, any similar instrument) to stand up from one criterion to another.

---

8. It will be noted that these values are based on a sample which is considerably smaller than the entire group dealt with; this results from the fact that the Personal Inventory was not administered during the entire period covered by our sample of records.
9. NDRC Project N-113, op. cit., P.iii.
10. The record cards supplied a psychiatrist's evaluations for a limited number of men who were referred to a psychiatrist for special interview. In general, the number of men falling in the four rubrics which the psychiatrist used in classifying these men is so small as not to warrant extended statistical analysis. Evaluations of some 200, however, who were classified as suffering from various degrees of "emotional instability", were correlated with the submarine officers' ratings on three of the rating-scale traits; none of the X^2's attain statistical significance. There is little in the data which would suggest that (for the present sample) psychiatrists' evaluations are any better predictors than the PI.

TABLE 4

Validity of the Personal Inventory in
Predicting Submarine Officers' Ratings of
347 Enlisted Men
(ATLANTIC FLEET)

Scale Traits	r*	M	σ
(1) Adaptability	-.033	2.56	.53
(2) Alertness	-.076	2.23	.63
(3) Courage	.009	2.26	.48
(4) Sociability	-.071	2.27	.52
(5) Leadership	-.045	1.88	.46
(6) Dependability	-.019	2.18	.52
(7) Overall Rating	.005	2.11	.46
Sum of Raw Ratings	-.048	15.49	2.64

TABLE 5

Validity of the Personal Inventory in
Predicting Submarine Officers' Ratings of
683 Enlisted Men
(PACIFIC FLEET)

Scale Traits	r*	M	σ
(1) Adaptability	-.061	2.67	.49
(2) Alertness	-.063	2.42	.60
(3) Courage	-.013	2.42	.53
(4) Sociability	-.037	2.35	.54
(5) Leadership	-.079	1.95	.53
(6) Dependability	-.037	2.39	.54
(7) Overall Rating	-.068	2.29	.57
Sum of Raw Ratings	-.071	16.48	2.76

* The signs of these coefficients should, of course, be negative since a low score on the Personal Inventory is a good one. None of these coefficients attains statistical significance.

Further, the reliability of the criterion[11] must be considered in evaluating all of the validity statistics presented in this Section. Unfortunately, the ratings used as criteria were made by only one officer, and as far as our data are concerned, only once. This, of course, makes it impossible for one to arrive at a direct statistical measure of the consistency of these evaluations. Evidence of an indirect sort, however, suggests that these ratings are reliable enough to permit significant correlations with other measures; achievement in some of the advanced schools, as measured by grades, did correlate with the rating data. This fact permits one to say that the low correlation between PI scores and ratings of performance can not entirely be explained by the unreliability of the criterion.

Evaluations reported elsewhere have demonstrated that the Personal Inventory is capable of making reliable discriminations among certain types of groups. The present statistics, however, do not indicate that it can predict ratings of the sort used in this study; unfortunately, with the limited amount of data available, it is not possible to definitely assign responsibility for the low cofrelations to the PI, the criterion, or to some factors left uncontrolled in the investigation.

B. The Validity of the Other Selection Measures and the School Grades

The other measures, --the Two-Hand Coordination Test, the Otis Self-Administering Test, the Confidential Questionnaire and the NRC Neurotic Inventory -- perform little better than the Personal Inventory (see Tables 6 and 7). Of the 32 validity coefficients reported for these four measures, only four approach statistical significance and these are hardly large enough to be of any predictive value. These results again, would lead one to suspect the criterion were it not for the fact, as was pointed out earlier, that certain of the school grades do correlate with the ratings.

Grades made in the Submarine Basic training courses can, to a certain extent, be used in predicting submarine officers' ratings. Likewise, certain of the Advanced School grades, in particular, those received after Diesel, Battery and Gyro, Ordnance, Sound, and Radio training, offer possibilities as predictors (see Tables 6 and 7). In general, the coefficients secured from the Pacific fleet data tend to be somewhat higher (and more of them attain statistical significance) than those from the Atlantic; this may suggest a slight superiority of the ratings received from the Pacific.

In comparing the validity coefficients which were obtained for the selection measures with those for the achievement measures, only one conclusion seems to be possible: if one accepts the criterion as valid, the existing achievement measures offer more possibilities as predictors of submarine performance. More extended work on both the criterion and the school grade offers more by way of possibilities than similar work with the existing selection battery.

---

11. Statistics, descriptive of the reliability of the Personal Inventory, are presented in the reports of NDRC Project N-113 --see, for example, OSRD Report No. 3963.

TABLE 6

Validity of Selection and Achievement Measures in Predicting Submarine Officers' Ratings of Enlisted Men - ATLANTIC FLEET

Test of Achievement Measure	Rating Scale Trait (1)	(3)	(7)	Sum	M	σ	N
Age	.143*	.068	.200*	.207*	20.01	3.04	360
Basic Training, Raw Grade	.164*	.034	.184*	.175*	3.20	.24	372
Basic Training, Class Standing	-.187*	-.140*	-.238*	-.246*	50.06	27.76	372
Diesel Training, Raw Guide	.322*	.199**	.262*	.351*	2.88	.29	107
Diesel Training, Class Standing	-.198**	-.138	-.249**	-.259*	52.38	28.30	107
Battery & Gyro, Raw Grade	-.121	.095	.058	.078	3.34	.24	79
Battery & Gyro, Class Standing	-.041	-.125	-.022	-.142	47.15	26.59	79
Ordnance Training, Raw Grade	.084	-.104	.178	.203	3.28	.13	43
Ordnance Training, Class Standing	-.092	.098	-.142	-.186	39.19	26.52	43
Sound Training, Raw Grade	.306**	.019	.199	.237	3.31	.18	51
Sound Training, Class Standing	-.256	.030	-.152	-.205	41.43	29.29	51
Radio Training, Raw Grade	-.063	.021	.000	.045	2.87	.26	31
Radio Training, Class Standing	.116	-.041	-.144	.010	49.45	25.73	31
Quartermaster, Raw Grade	-.126	-.125	-.239	-.168	3.51	.15	39
Quartermaster, Class Standing	.036	.121	.196	.068	40.00	21.82	39
Lookout Training, Raw Grade	-.057	-.027	.037	.049	3.35	.25	43
Lookout Training, Class Standing	.136	.017	-.140	-.106	42.37	29.18	43
Tank Performance	.127	.098	.121	.031	2.86	.66	495
Two-Hand Coordination Test	.038	.078	.028	.080	94.99	29.12	369
Otis Self-Administering Test	.092	.063	.038	.097	104.80	9.64	372
New London Confidential Question	-.008	.009	.003	-.017	-9.56	5.09	324
NRC Neurotic Inventory	-.001	-.074	.023	-.020	12.19	3.31	298

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

TABLE 7

Validity of Selection and Achievement Measures in Predicting Submarine Officers' Ratings of Enlisted Men - PACIFIC FLEET

Test of Achievement Measure	Rating Scale Trait (1)	(3)	(7)	Sum	M	σ	N
Age	.125*	.117*	.166*	.174*	20.30	2.87	751
Basic Training, Raw Grade	.169*	.159*	.183*	.246*	3.29	.22	797
Basic Training, Class Standing	-.166*	-.082**	-.183*	-.209*	46.00	28.77	797
Diesel Training, Raw Grade	.131	.058	.058	.107	3.05	.25	191
Diesel Training, Class Standing	-.051	.031	-.070	-.070	46.60	29.16	191
Battery & Gyro, Raw Grade	.299*	.230*	.328*	.343*	3.37	.28	179
Battery & Gyro, Class Standing	-.299*	-.214*	-.258*	-.289*	46.02	28.38	179
Ordnance Training, Raw Grade	.156**	.287*	.282*	.351*	3.33	.15	194
Ordnance Training, Class Standing	-.097	-.159**	-.206*	-.229*	48.44	28.93	194
Sound Training, Raw Grade	.160	.135	.389*	.317*	3.28	.19	114
Sound Training, Class Standing	-.146	-.152	-.327*	-.281*	49.70	28.74	114
Radio Training, Raw Grade	.242	.203	.388*	.357*	2.88	.36	56
Radio Training, Class Standing	-.292**	-.226	-.399*	-.367*	52.23	28.14	56
Quartermaster, Raw Grade	.176	-.012	-.203	.087	3.44	.22	59
Quartermaster, Class Standing	-.145	-.095	-.145	.073	46.20	27.47	59
Lookout Training, Raw Grade	.191	-.196	.139	.093	3.30	.25	30
Lookout Training, Class Standing	-.078	.219	-.142	-.035	46.70	25.69	30
Tank Performance	.122	.121	.099	.057	2.93	.68	806
Two-Hand Coordination Test	.084**	.073	.083**	.110**	94.27	27.69	746
Otis-Self-Administering Test	.030	-.012	.036	.048	103.16	10.05	796
New London Confidential Question	-.061	.020	-.010	-.013	-8.65	4.91	532
NRC Neurotic Inventory	-.034	-.041	-.120**	-.095	12.24	3.43	443

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

VI. SUMMARY

The record cards of the Medical Research Laboratory at the New London Submarine Base supplied the test scores and school grades of some 1700 men who were later evaluated by submarine officers on a seven-trait graphic rating scale. The present report attempts to evaluate the test scores and school grades as predictors of later submarine performance.

In making this study, the following procedures were carried out:

1. The submarine ratings were subjected to extended analysis with the end in view of selecting (a) those vessels whose ratings offered the greatest possibilities as criteria, and (b) those traits which best represented the abilities and performances measured by the scale as a whole. Ratings on three of the scale traits, (Adaptability, Courage and Adequacy of Overall Performance) along with the sum of the ratings for all seven traits, from 29 submarines of the Atlantic Fleet and 41 from the Pacific were finally accepted for use as criteria.

2. Test scores and school grades were correlated with the ratings on the selected traits. Atlantic and Pacific fleets were treated independently.

The results of the validity analysis may be summarized as follows:

1. Scores on the Personal Inventory, and on the other selection measures employed at the New London Submarine Base, were found to be practically unrelated to performance as it is described by officers' ratings. The validity coefficients are probably greatly attenuated, since the criteria are of questionable reliability. Further, the men are selected with respect to the abilities investigated here; and the criteria may be considered as inappropriate, in the sense that the items which comprise the selection measures were chosen for their ability to make discriminations of the type which may not at all be related to those called for by the present criteria.

2. An analysis of psychiatrists' evaluations of men judged to be "emotionally unstable" suggests that, for this sample at least, these evaluations are no more valid than those made by the paper and pencil tests of the selection battery.

3. On the other hand, grades made in the submarine basic training courses can, to a certain extent, be used in predicting ratings made by submarine officers. In a similar manner, certain of the advanced school grades -- those received after Diesel, Battery and Gyro, Ordnance, Sound, and Radio training -- also offer possibilities as predictors.

4. School grades seem to offer the most by way of possibilities for improved prediction. For this reason, it would seem most reasonable to direct further work toward the improvement of both the school grade and the criterion.

BLANK PAGE