

2-23

AD 663861

Project 44  
Division 7, NDRC

Contract  
OEMsr-570

A REPORT OF  
RESEARCH ON SELECTION TESTS  
AT THE  
U. S. SUBMARINE BASE  
NEW LONDON, CONNECTICUT

OSRD REPORT  
No. 1770

Brown  
University

Reproduced by the  
**CLEARINGHOUSE**  
for Federal Scientific & Technical  
Information Springfield Va. 22151

This document has been approved  
for public release and sale; its  
distribution is unlimited.

DDC  
RECEIVED  
JAN 19 1968  
RECORDED

CB

72

AD 663861

~~Restricted~~

PROJECT: RESEARCH ON THE SELECTION OF  
SERVICE PERSONNEL, WITH SPECIAL REFERENCE  
TO EMOTIONAL FACTORS

Project 44, Division 7

National Defense Research Committee

June 28, 1943

A REPORT OF RESEARCH ON SELECTION TESTS  
AT THE U. S. SUBMARINE BASE, NEW LONDON

by

C. H. Graham, L. A. Riggs, N. R. Bartlett,  
R. N. Berry, W. S. Verplanck, R. L. Solomon  
and C. G. Mueller

Contract No. OEMsr-570  
of the  
Office of Scientific Research and Development

OSRD Report No. 1770

Copy No. 128 of 150

Brown University  
Providence, R.I.

~~RESTRICTED~~

~~RECEIVED~~

At the Submarine Base we have felt that the selection of "psychologically fit" men for submarine service constitutes a most important problem. Because of our recognition of the need for improved selection methods, we have collaborated in the work of Project 44 with enthusiastic interest. We believe that the results may prove valuable in a field where contributions to method are of fundamental importance.

  
C. W. Shilling

Commander (MC) USN  
Medical Officer in Charge  
Medical Research Laboratory  
U. S. Submarine Base  
New London, Connecticut

~~RECEIVED~~

~~CONFIDENTIAL~~

The Staff of Project 44 expresses appreciation to the following Naval Officers and their staffs for active collaboration in the research described in this report:

- Commander C. W. Shilling (MC) USN
- Commander T. E. Willmon (MC) USN
- Lieutenant-Commander Calvert Stein (MC) USNR
- Lieutenant (jg) I. A. Everley (HC) USN
- Ensign J. L. Blizzard (HC) USN  
all of the Staff of the Medical Research  
Laboratory;
- Commander J. W. Blanchard, USN  
of the Submarine School; and
- Lieutenant G. W. Albin, USNR  
of the Submarine Escape Training Tank.

Special thanks are due Captain E. F. Cutts,  
Commander, Submarine Base, New London,  
for his provision of space and facilities.

~~CONFIDENTIAL~~

## Table of Contents

<u>SECTION</u>	<u>PAGE</u>
Recommendations	iv
Summary	iv
Introduction	1
The background of the research	2
The tests	5
Test intercorrelations and reliabilities	7
Results: Psychiatric criterion	7
Results: Submarine men tested after patrol	16
Results: Tank performance	21
Discussion	23
Appendix A: The shock-fusion test	A1
Appendix B: Item analysis of the NRC Neurotic Inventory	B1
Appendix C: The New London-NDRC Confidential Questionnaire	C1
Appendix D: Scoring Sheet for the New London-NDRC Confidential Questionnaire	D1
Appendix E: Standardization of the New London-NDRC Confidential Questionnaire	E1
Appendix F: Officers' Rating Scale	F1
Appendix G: Correlations of the first 12 items in the Officers' Rating Scale with the final item	G1

~~CONFIDENTIAL~~

RECOMMENDATIONS

The present report demonstrates that batteries combining three tests, an intelligence test, a personal inventory, and a 2-hand coordination test, relate significantly to (a) psychiatrists' judgments of "psychological fitness" of entering Submarine School candidates, and (b) officers' judgments of experienced submarine men. Because of the significant relationship between test scores and psychiatrists' and officers' judgments, the test batteries may prove valuable as aids in eliminating "psychologically unfit" individuals from entrance to the Submarine School. It is possible that similar applications might be made in other Service branches after due consideration of personnel requirements.

The following recommendations are based upon the experimental results:

1. It is recommended that some of the procedures, for which results are given in this report, be adopted as part of the methods of the submarine medical examination.

2. Two alternative applications, discussed on pages 23 to 25, might be made at the Submarine Base; both relate to procedures discussed on pages 10 and 11.

(a) One method involves a "selection cost" of about 33 per cent for a decrease of about 70 per cent in the number of "unfit" men.

(b) The other involves little "selection cost" and requires psychiatric examination only for those men among the entering candidates who perform most poorly on the tests (suggested: the lowest 30 per cent).

Procedure (b) is recommended.

3. Advantages of procedure (b), as it might be applied at the Submarine Base, are discussed on page 24.

4. Procedure (a) might have advantages for some Service branches where "selection cost" is of secondary importance to quality of personnel. Procedure (b) might prove significant as an effective aid to Service psychiatrists.

5. It is recommended that an experimental program, directed to further refinements in procedures, be maintained at the Submarine Base.

SUMMARY

1. This report presents results of research conducted at the New London Submarine Base. The research is concerned with the development of tests for identifying men who are "psychologically unfit".

2. Three criteria of "unfitness" have been adopted:

- (a) Psychiatric judgment,
- (b) Officers' judgments of men  
from operating crews of  
submarines, and
- (c) Escape Tank performance.

~~CONFIDENTIAL~~

It is hoped that a later report will deal with officers' judgments of men, previously tested, after the men have seen service at sea.

3. Tests used were the following: The NRC Inventory, the New London-NDRC Confidential Questionnaire, the Personal Inventory, and a 2-hand coordination test. Data on Otis intelligence test scores have also been analyzed. Our treatment considers results obtained from three batteries, each made up of the 2-hand test and the Otis in combination with one of the inventories. The name of the included inventory gives the title to the battery.
4. All of the individual tests relate significantly to psychiatric criterion (judgment in terms of a "Good"- "Bad" dichotomy).
5. Individual tests, combined into the New London and Personal Inventory batteries according to the procedure described on pages 10 and 11, predict the psychiatric dichotomy in highly significant degree and to an extent which improves considerably on results obtained by the intelligence test alone.
6. A battery involving items of the New London Inventory relates significantly to officers' judgments of men from operating crews of submarines (low ratings vs. high ratings). The New London battery is more discriminative than the Otis test alone. The battery containing the NRC Inventory is probably not so discriminative as the New London battery in this series of tests.
7. Tank performance records were examined for all of the men in our experimental group who were allowed to take Tank training after our tests. Analyses were made with respect to "failures" and "successes" in Tank performance for the NRC battery. Failure involves certain symptoms of "panic", etc., which are discussed on page 21. At cut-offs IV and V of Figure 3, the NRC battery discriminates "failures" from "successes" in significant degree for a population of men whose data entered into test standardization. The battery failed to show significant discrimination for a non-standardization group. Further analyses of Tank performance data may be valuable.
8. Figure 4 gives, for psychiatric criterion, a summary of selection costs and gains. If the manpower pool is large enough to provide 141 men for testing, the number of "bad" men in a class of 100 is reduced from 15 to 4 (New London Battery). Supporting data are given by the other batteries.
9. Figure 5 gives data on officers' judgments of men from operating crews of submarines. For the New London battery, a pool of 125-135 men leads to a crew of 100 men containing 5 or 6 "bad" men after testing, as contrasted with a crew containing 13 "bad" men when no tests are applied.
10. Two specific test procedures, applicable for use in the submarine medical examination, are discussed on pages 23 to 25.
  - (a) One method involves a selection cost of about 33 per cent for a decrease of about 70 per cent in the number of "Bads".
  - (b) The other method involves little selection cost and requires psychiatric examination only for those men among the entering candidates who perform most poorly on the tests.
11. An evaluation of the research, together with recommendations for future work, is given on page 23.

~~CONFIDENTIAL~~

## INTRODUCTION

This report presents an account of research conducted by us from about June 1, 1942 to May 1, 1943, at the United States Submarine Base in New London. The research has been aimed at the development of tests for identifying men who are "psychologically unfit"<sup>1</sup> as regards "temperamental characteristics".

The present report must be considered a progress report. It records results based on three criteria of "unfitness": (1) psychiatric judgment, (2) officers' judgments of men who have undergone patrol duty prior to testing, and (3) Escape Tank performance. Some of our best validation will be obtained when we receive officers' ratings of men, previously tested at New London, after the men have seen service at sea. We hope that these data will be available in sufficient numbers for a later analysis. Up to the present, they have arrived in too few numbers to merit inclusion in the present account.

Psychiatric judgment serves as a convenient basis for an approach to the problem of "psychological unfitness". Unfortunately, from an experimental viewpoint it has the limitation that, at New London, the psychiatrist usually knows the subject's intelligence score at the time of making a rating.

The extent to which this knowledge enters into a "total" judgment is uncertain. Because of this uncertainty and because of the fact that our research is primarily directed to a study of "temperamental characteristics", we have considered only those men who are listed as emotionally unstable, psychoneurotic or neurologically abnormal for inclusion in our criterion group. Certain of these men are judged to have low intelligence, but no man rejected solely on the basis of low intelligence is included in our research populations. (However, we have kept a full account of each man in the latter category; see footnote 1.)

Despite the fact that our research has not considered men who were rejected for low intelligence alone, we have found it necessary to analyze intelligence test data. The analysis shows that intelligence test score is significantly related to psychiatric judgment of temperamental characteristics, even when men disqualified solely on the basis of low intelligence are eliminated from our populations. This result may mean that psychiatric judgment of temperament is influenced by intelligence test score, or that intelligence is an inescapable component of the "total" personality picture. Discussion of this point is probably fruitless. The fact of the correlation is the important thing.

On the basis of the psychiatric criterion, we have examined data from three "inventories" and a 2-hand coordination test. The analysis demonstrates that it is possible to construct test batteries of three tests (the 3-test batteries being made up of different combinations of 5 tests) which correlate with psychiatric opinion better than does the intelligence test alone.

---

<sup>1</sup>Different tests have been given at different periods in our research, a few men did not complete all of the tests constituting a given battery, and some tests required standardization. Because of all of these factors, the numbers of men in the populations for the various tests and batteries vary.

Not included in our populations are 204 men who were tested but received no psychiatric examination; and 167 men who were tested but were later disqualified (on the basis of physical [N=97] and intellectual deficiencies [N=70]) from submarine service.

~~CONFIDENTIAL~~

The approach to the problem of "psychological unfitness" by way of psychiatric judgment was chosen because of its convenience and the speed with which fruitful preliminary data might be obtained. Once the tests had been analyzed on the basis of psychiatric judgment, they were applied in situations involving new criteria: (a) officers' ratings of men who had undergone submarine duty, and (b) Escape Tank performance (described below). Neither criterion is subject to the objection that intelligence test score may have influenced validating performance. On the other hand, each suffers from the fact that no men disqualified on entering the Submarine School are available for study.

Unfortunately, because of the necessity for developing and improving "inventories" and the small population of experienced submarine men available, data on Tank performance and officers' ratings are not as satisfactory as they are for psychiatric criterion. Within their limits, however, the data from officers' judgments provide supporting evidence for the usable nature of the tests. More experiments are required before definitive conclusions may be reached with regard to Tank performance.

#### THE BACKGROUND OF THE RESEARCH

The New London Submarine Base presents a number of advantages for conducting a research program. Here is located the only Submarine School for enlisted personnel. More than 90 percent of enlisted submarine personnel are graduates of this school, and so one can examine at this Base a good cross-section of the men reporting for submarine training. Furthermore, the submarine service itself is a relatively small and well-knit organization so that there is possible access to most data or records that one might wish to use. Finally, there are available at the Base, or within the submarine service, several valuable criteria for validating and standardizing personnel tests.

Requisites for admission of enlisted personnel to the Submarine School are (1) medical certification after a physical and psychological examination and (2) successful negotiation of two escapes with the submarine escape appliance from a depth of at least 18 feet of water. During the time of this research program, approximately 85 percent of the candidates met these requirements and were admitted to Submarine School. The population of candidates is drawn from the whole Navy: training stations, training schools, the Fleet, and all activities. All rates, from Apprentice Seaman to First Class Petty Officer, are well represented. Theoretically, all men are volunteers for submarines, but actually a few report contrary to their own wishes.

Our research program was carried out on a large random sample of the population of candidates. During the first six weeks of the program (until July 15, 1942) various methods were used to obtain subjects. Statistical study showed the sample drawn in those six weeks to be a random one of the then current population with regard to psychiatric ratings, school grades, physical disabilities, and Tank performance. After mid-July only one method was used. After that date, the men tested each day were drawn at random from the population reporting on that particular day for medical examinations.

In addition to this random sample group, two other groups have been examined systematically. The first additional group consists of men (N=29) referred by the crew of the Escape Training Tank because of manifestations of emotional disturbance. The second group (N=161) is made up of samples of

operating crews of submarines that, for one reason or another, have put in at the New London Base.

Our populations contain no ranks higher than Petty Officer, First Class, and almost no colored men. Ages range from 17 to 35, the ages being typically in the late teens or early twenties. Length of service seldom exceeded one year; indeed, the major part of the candidates came straight from training schools and had had no sea experience.

The Otis S-A Test of Mental Ability has long been in routine use in the Submarine Medical Examination and plays an important role in selective procedures. The Otis was always used as the first in the sequence of tests making up the Project 44 battery. The second test was always the NRC Form R-2. The third test was the 2-hand coordination device. All men, therefore, had at least the minimum battery consisting of NRC Form R-2 and the 2-hand coordination test. In addition, intelligence test score was entered in the records of each man.

Other tests have been investigated at various times, each one except (2) following, being applied to a sample of several hundred subjects. The additional tests that have been used are: (1) the Willoughby Personality Inventory; (2) the Bennett Test of Mechanical Comprehension; (3) the Personal Inventory; (4) the New London-NDRC Confidential Questionnaire; and (5) the Shock-Fusion Reaction-Time test. All of these additional tests, except the last-named, are group tests. The last is an individual test.

For the population of candidates appearing for medical examination certain information is available.

For one thing, nearly every man is interviewed for from five to twenty minutes by a psychiatrist. Slightly less than half of the sample tested each day were interviewed by the psychiatrist prior to administration of the battery. The balance were interviewed subsequent to completion of the battery. In the majority of the interviews (at least 90 percent) the Otis score was known by the psychiatrist, but no other test scores were ever made available to him. Except insofar as the Otis may have influenced his classifications, the psychiatrist operated entirely independently of the Project 44 battery. On the other hand, psychiatric classifications were copied from the psychiatrist's office only after the individuals concerned had completed the battery. Thus there is usually a psychiatric classification available for each candidate, and this classification is independent of all but the intelligence test.

At the conclusion of the series of medical tests, each man is interviewed briefly by the senior medical officer in charge of submarine examining. The latter, largely on the basis of all the physical and psychological data recorded that day on the candidate's card, renders a decision whether the candidate is qualified medically for submarine service. No tests of the Project 44 battery were ever recorded on the candidate's card, and thus the senior medical officer could have been influenced only by psychiatric judgment and intelligence score.

Ordinarily, within ten days following the medical examination, the qualified candidates are sent to the Submarine Escape Training Tank for training in the use of the submarine escape appliance ("lung"). This training requires 2 to 3 hours, although the time spent per man in actual performance is less than an hour. During the course of this training, a small percentage

are disqualified because of inability to withstand exposure to high atmospheric pressure. Again, about one percent manifest abnormal postures, inability to follow instructions, excessive trembling, or other responses so unusual that the Tank instructors feel it unsafe to continue underwater training. Sometimes, eccentric behavior is not so markedly unusual, and, in the opinion of the instructors, training may be safely continued. In such cases, the individual is carried throughout the entire course of training. But all bizarre or eccentric cases are brought to the attention of the medical officer at the Tank, who determines in each instance from the instructor's account whether the man will be "Disqualified temperamentally". During the period of this research, all eccentric cases were referred by the officer-in-charge of the tank to the Brown unit for study, and if they had not already been tested as members of the sample group, were given the battery. Twenty-nine referrals were made to us of men who had not been previously tested in the representative population.

Every two weeks a new class in the basic submarine course is started. With few exceptions the men who are medically qualified and who have successfully completed the required minimum of Tank training begin school in the next succeeding class.

The men completing the basic six-weeks' course are graded by their section leaders on a complex basis including notebooks, brief examinations, attentiveness in class, aptitude demonstrated on school cruises, dependability, and general deportment. Each section leader (an experienced submarine Chief Petty Officer) is responsible for the grades of his section of fourteen men. Each student is then ranked on the basis of his standing in the entire class, and is assigned, in addition, both a percentage and a Navy grade. The Navy grade is on a scale from 0.0 to 4.0.

Following graduation from the basic course, men are assigned either to some submarine division or to advanced submarine schools for specialization in their particular rates. A great majority are assigned to advanced schools.

Arrangement was made for insertion of blank performance reports inside the jacket of the service records of all men tested. Insertion of the blank reports has been accomplished for approximately 90 percent of all the individuals tested by the Project 44 staff. On the outside of the jacket a small notice is pasted, "Important Comsublant Directive enclosed". The directive requests the commanding officer to complete and return the performance report at the end of the first war patrol made by the individual concerned. As an estimate, return of reports may be expected approximately nine months from the date of test. This estimate is based on analysis of assignments of sample submarine classes following graduation from the basic submarine course. These reports should afford some information concerning the individual's adjustment to the strenuous duty of submarine war patrols.

In summary, then, for validation the following information was sought out and as it has become available, has been tabulated for the men tested from the entering School population and for the tank referrals: (1) medical (including psychiatric) classification, (2) Tank performance, and (3) service performance in submarines.

For the additional groups of subjects only a part of this information was, of course, available as the derivation of the groups indicates. For example,

the only validation data obtained from the populations of operating submarine crews were officers' opinions of previous service performance.

### THE TESTS

Tests used in the course of the present experiments were: (a) the NRC Neurotic Inventory (Form R-2); (b) the New London-NDRC Confidential Questionnaire; (c) the Personal Inventory; and (d) the 2-hand coordination test. The Otis S-A Test of Mental Ability was in routine use in the Submarine School medical examination, and we have analyzed the correlations of this test with our criteria.

The NRC Neurotic Inventory (Form R-2) was developed by a Committee of the National Research Council and is an "inventory" emphasizing attitudinal factors. The score sheet is adapted to machine scoring. In the present research we have used a scoring system based on our own item analysis. Results of the item analysis are described in Appendix B.

The New London-NDRC Confidential Questionnaire was developed and standardized on a Submarine School population with a criterion of psychiatric judgment and Tank failure. The test places emphasis on biographical items. Appendix C contains the Questionnaire and Appendix D, the scoring sheet. Appendix E gives an account of standardization.

The Personal Inventory was developed under contract OEMsr-834 of the Office of Scientific Research and Development. The contract was recommended by the Committee on Service Personnel--Selection and Training, of the National Research Council. Walter C. Shipley is Project Director. Emphasis is placed upon biographical items, and the test has been standardized on "normal" personnel and "psychiatric cases" at the United States Naval Training Station, Newport, Rhode Island. Our analyses at New London have been based upon the scoring method devised in the Newport study (see OSRD Report No. 1606).

The 2-hand coordination device is shown in Figure 1. This apparatus was built by Carl Heinrich of Boston, Massachusetts. It presents the subject with a motor pursuit task in which the moving target is a small disc. The disc is supported by a lever which is driven by an electric motor around the periphery of an irregularly shaped cam. The subject is instructed to keep a small button in continuous contact with the irregularly moving disc. The button is positioned by two hand cranks and screws at right angles to each other. The subject is scored on his ability to manipulate the hand cranks in such a way as to keep the button in contact with the disc. After each run, the score is read from the dial of an electrically operated stop clock which measures the total amount of time during which contact was maintained. Since each run consumes one minute, and since the clock reads in units of 0.01 second, the dial reading represents the percentage of time the subject maintained contact during a given course. Low scores indicate poor pursuit performance, and high scores, good.

Tests in addition to those mentioned above have been used at various times.

Preliminary work before the start of the New London program showed that (a) a steadiness test, (b) an eye-hand coordination pursuit test, and (c) a test of immediate memory do not differentiate undesirable from desirable men.

The following tests have been given at New London in addition to the battery finally adopted: (a) the Willoughby Personality Inventory, (b) the Bennett Test of Mechanical Comprehension, and (c) the Shock-Fusion Reaction Time Test.

The Willoughby Personality Inventory lacks discrimination. The Bennett Test was given only to a few men.

The Shock-Fusion Test has given good results. It is described in Appendix A. However, it is an individual test, and its administration requires such a long time (45 minutes), that we considered it unsuitable as a routine selection device.

#### TEST INTERCORRELATIONS AND RELIABILITIES

Table I gives data on the intercorrelations (product-moment) existing among the tests. The numbers of men in the populations on which the intercorrelations were determined are indicated by the numbers in parentheses. In all cases populations are random samples from groups not used for standardization purposes. Only 90 men took both the Personal Inventory and the New London-NDRC Confidential Questionnaire.

It may be observed that all of the intercorrelations are of low order. The 2-hand coordination test correlates only in a slight degree with any of the other tests, including the intelligence test. The intelligence test does not correlate highly with any other test. The New London Personal Questionnaire and the NRC Inventory correlates to the same significant amount with the Personal Inventory. This may mean that the Personal Inventory has characteristics common to both of the other inventories.

No test correlates with any other test to such a degree as to cause exclusion of one test from a battery because of the presence of another test.

Split-half reliabilities, extended by the factor  $2r/1+r$ , are as follows: 2-hand, .92; NRC, .54; New London, .75; Personal Inventory, .77.

Data for the 2-hand (trials 2 and 4 vs. 3 and 5 in a 5-trial series) and NRC tests were obtained from the records of the same 400 men, arranged to give a representative sample with regard to psychiatric criterion. Data for the New London test were obtained in the same manner from another group of 400. No standardizing cases are involved in the computations for any test. The figure for the Personal Inventory is taken from OSRD Report No. 1606 (Newport population).

#### RESULTS: PSYCHIATRIC CRITERION

##### Data on individual tests

The populations on which we present data for the individual tests do not involve any cases on which the tests were standardized.

Results showing the degree of correspondence between individual test scores and psychiatric judgment are presented in Table II. Table II gives, for different cut-offs, the percentages in the "Good" and "Bad" populations failing each of the tests, the percentage of failures in the total population,

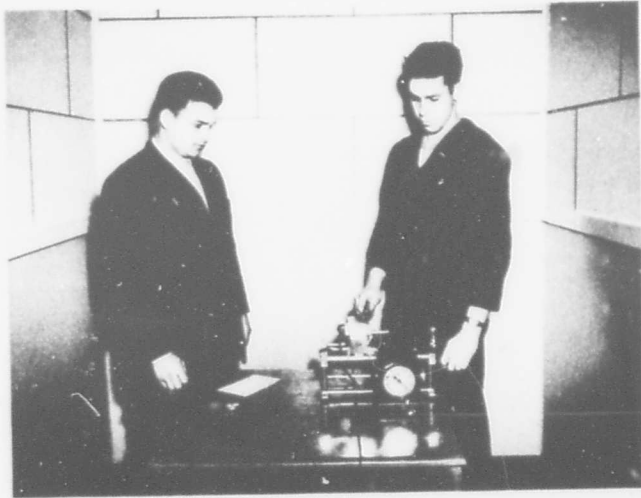


Figure 1  
The two-hand coordination apparatus.

TABLE I

## Intercorrelations Among the Tests

Number in parentheses indicates number in group used for computation.

	NRC	2-Hand	Otis	New London	Personal Inventory
NRC		-.08 (500)	-.27 (500)	.22 (500)	.38 (400)
2-Hand	-.08 (500)		.21 (500)	-.06 (500)	-.15 (400)
Otis	-.27 (500)	.21 (500)		-.11 (500)	-.17 (400)
New London	.22 (500)	-.06 (500)	-.11 (500)		.36 (90)
Personal Inventory	.38 (400)	-.15 (400)	-.17 (400)	.36 (90)	

TABLE II

## EFFECTIVENESS OF INDIVIDUAL TESTS IN SELECTION

TEST	Cut-off Score	Number "Goods" Failing	Percent "Goods" Failing	Number "Bads" Failing	Percent "Bads" Failing	Of the	Percent of Total Population Failing	Chi square	r
						Number Failing, This Percentage Is "Bad"			
Two-Hand Coordination (N=1921)	15	3	0.18	2	0.7	40	0.3	2.2	.47
	25	45	2.8	21	7.1	32	3.4	14.0	.37
	30	98	6.0	36	12.1	27	7.0	14.4	.29
	35	184	11.3	55	18.5	23	12.4	11.9	.22
	40	331	20.4	90	30.3	21	21.9	14.4	.20
	42	400	24.6	111	37.4	22	26.6	20.9	.23
	45	543	33.4	136	45.8	20	35.3	16.7	.20
	50	780	48.0	177	59.6	18	49.8	13.4	.18
Otis S-A Intelligence (N=1940)	74	5	0.3	12	4.0	71	0.9	42.3	.79
	80	22	1.3	45	15.2	67	3.5	144.2	.78
	86	111	6.8	115	38.8	51	11.7	250.3	.70
	89	188	11.4	154	52.0	45	17.6	283.9	.69
	92	318	19.3	179	60.5	36	25.6	172.9	.63
	95	475	28.9	204	68.9	30	35.0	176.3	.59
	101	818	49.8	244	82.4	23	54.7	107.9	.55
	104	1012	61.6	258	87.2	20	65.5	72.7	.52
NRC Neurotic Inventory (N=1310)	21	16	1.5	17	8.2	52	2.5	32.6	.64
	20	28	2.5	28	13.5	50	4.3	50.9	.61
	19	49	4.4	41	19.7	46	6.9	63.6	.58
	18	75	6.8	50	24.0	40	9.5	60.5	.53
	17	109	9.9	69	33.2	39	13.6	80.6	.54
	16	184	16.7	88	42.3	32	20.8	69.7	.47
	15	246	24.0	111	53.4	31	27.3	74.2	.47
New London NDRC Personal Questionnaire (N=525)	+6	6	1.4	12	14.5	67	3.4	32.0	.75
	+5	8	1.8	15	18.2	65	4.4	39.8	.74
	+4	18	4.1	17	20.6	49	6.7	26.9	.59
	+3	30	6.8	20	24.2	40	9.5	20.7	.49
	+2	39	8.8	21	25.4	35	11.4	15.6	.42
	+1	52	11.7	25	30.3	32	14.7	17.9	.42
	0	68	15.3	30	36.3	31	18.8	18.5	.40
	-1	90	20.3	37	44.8	29	24.2	20.1	.40
Personal Inventory (N=522)	24	2	0.4	9	13.6	82	2.1	48.3	.90
	22	7	1.5	14	21.2	67	4.0	56.3	.82
	20	15	3.3	15	22.7	50	5.7	40.1	.70
	18	23	5.0	20	30.3	47	8.2	49.2	.69
	16	37	8.1	21	31.8	36	11.1	33.1	.58
	14	75	16.4	31	47.0	29	20.3	33.2	.53
	12	119	26.1	37	56.1	24	29.9	24.8	.47

and the percentage of failures constituted by "Bads".

All tests show a relationship with criterion significantly above chance, and the relationship, for each test, is consistent for all cut-off values.

An indication of the absolute effectiveness of the individual tests, in terms of criterion, is given by the values of chi square and the tetrachoric correlations between test scores and psychiatric judgment. Discrimination indicated by chi square is, in all cases except the 2-hand test at cut-off 15, highly significant at levels of  $p$  less than .01.

Table III presents certain facts of Table II in such a manner that essential findings are summarized. Tetrachoric correlations have been calculated at cut-offs of the individual tests which fail nearly comparable percentages of the total population taking each test. For example, the tetrachoric correlations with criterion, calculated at cut-offs which fail more than 24 percent and less than 30 percent of the test populations, are as follows: 2-hand, .23; Otis, .63; NRC, .47; New London, .40; and Personal Inventory, .57.

TABLE III

Tetrachoric correlations at cut-offs giving comparable failure rates for the individual tests

<u>Percentage Failed</u>	<u>2-hand</u>	<u>Otis</u>	<u>NRC</u>	<u>New London</u>	<u>Personal Inventory</u>
7 - 12	.29	.70	.53	.42	.71
12 - 18	.22	.69	.54	.42	.61
24 - 30	.23	.63	.47	.40	.57
Average of 3 levels	.25	.67	.51	.41	.63

At all levels of discrimination the Personal Inventory and the Otis seem to be superior to the rest of the battery in identifying the "Bads". The 2-hand coordination test is our least effective test.

Method for combining individual tests in batteries

Combining tests to give satisfactory rates of rejection under the conditions of our experiment involves an arbitrary methodology.<sup>2</sup>

A first step involves the application of the 2-hand test. All men who have a score below 25 (average of 5 trials) are disqualified. At this cut-off level the 2-hand test has good validity, and the odds in favor of the identification of a "bad" man are better than they are at higher cut-offs. In

---

2

We are indebted to Lieutenant-Commander J. W. Dunlap, USNR, for a discussion of the method.

addition, the test appeals to the candidate's interest, thus enhancing rapport. As a test it "makes sense" to the candidate.

A second step requires the setting up of a grid with "inventory" scores marked off one axis and Otis scores on the other. Figure 2 demonstrates this type of construction.

Neither "inventory" nor Otis step intervals must be constant, although they may be. In our particular case the choice of step intervals was governed by three considerations: (1) Our previous knowledge of the efficiency of various cut-off scores for individual tests. Class intervals were chosen so that previous interval cut-off systems could be represented by the column-row arrangement. (2) The need for fine gradations at the "bad" end of the distribution and less fine gradations at the "good". (3) The necessity for having enough cases in each category to make statistical procedures meaningful.

Because of these considerations, the step intervals used are not constant in magnitude, but have been based upon our impressions as to what might constitute fruitful categories.

Three grids were laid out, one for the Personal Inventory - Otis combination, one for the New London-Otis combination, and one for the NRC - Otis combination. The number of "Bads" as a percentage of the total number of cases lying in each cell was determined, and this figure is represented in Figure 2 as the height of the block based on each cell. It will be obvious in Figure 2 that the "poor" scores converge on the farthest corner of each grid, while the "good" scores converge on the nearest corner. It is seen that, for all "inventory" - Otis combinations, the heights of the blocks generally increase as the cells progress toward the poor scores and decrease as the cells progress toward the good scores. This means that the "Bads" become a greater and greater percentage of the "failed" population as we approach the "poor" score corner of the grid.

Figure 3 gives a representation of the grid system which is applicable to any of the Otis - "inventory" combinations. The grid cells are marked arbitrarily with Roman numerals. Thus an individual with an Otis score lying anywhere between 0 and 104 will be represented by Roman numerals I, II, III, IV, or V, depending upon his scores in the "inventory". For example, a man with an Otis score of 89 and a Personal Inventory score of 17 is represented by a cell containing the numeral II.

A given cut-off level is represented by the cumulated cells in a given category plus all those of a lower category. Thus, in Figure 3 all those cases in the squares which contain the numeral I or who fail the 2-hand test are considered to constitute the lowest cut-off group. The next higher cut-off group consists of those cases whose scores fall in the squares containing the numeral II plus all those containing the numeral I. The next higher group contains those cases falling in the III squares plus all those falling in the I's and II's. The highest cut-off category involves all of the squares containing Roman numerals.

#### A comparison of individual tests and batteries

Table IV gives complete data at various cut-off areas for the 3 batteries of tests employed in the present experiments. Discrimination indicated by

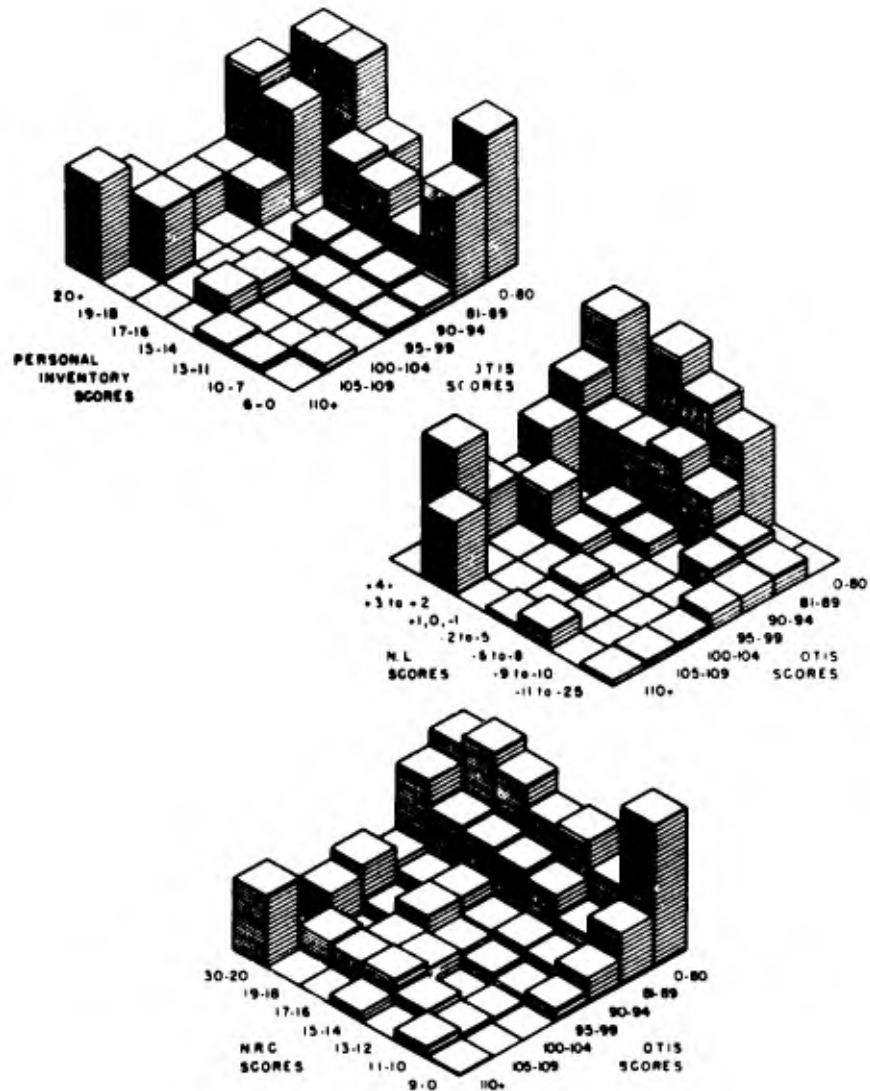


FIGURE 2. "Bads" as percentage of men in each combination test score category. The categories refer to "inventory"-Otis score combinations. Men failing the 2-hand test are eliminated prior to this step. The scores for each test are divided into seven arbitrary categories. The categories are arranged on a square grid so that all possible combinations of scores are obtained. The "poor" scores converge on the farthest corner of each grid, while the "good" scores converge on the nearest corner. The height of each block on the grid indicates the "Bads" as a percentage of all the men ("Goods" plus "Bads") falling in that block. It is seen that, for all "inventory"-Otis combinations, the heights of the blocks generally increase as the blocks progress toward the poor scores and decrease as the cells progress toward the good scores.

The results of this treatment are given for each of the three "inventory" batteries: NRC, Personal Inventory, and New London Inventory. In each case the categories of Otis scores are set against the categories for the scores of the respective Inventories.

NL	20+4	3+2	1+1	2+5	6+8	9+10	11+25
PI	20+	19-18	17-16	15-14	13-11	10-7	6-0
NRC	3020	19-18	17-16	15-14	13-12	11-10	9-0
0-80	I	I	I	I	I	II	II
81-89	I	II	II	III	III	III	III
90-94	I	II	IV	IV	IV	V	V
95-99	I	III	IV				
100-104	I	III	IV				
105-109	II	III	V				
110+	II	III	V				

FIGURE 3. Proposed system of cut-offs for screening men with the combination test score procedure. Otis categories are set against each of the "inventory" categories. Suitable cut-offs are outlined, taken on the basis of results such as those leading to Figure 2. Area I, containing the cells marked I, is the area failing the lowest number of men. The areas are taken cumulatively; that is, when area III is used, all men whose scores place them in area I, II, or III are eliminated by the battery. The scores for all inventories are given so that only one diagram is necessary.

T A B L E I V

RESULTS FOR THREE BATTERIES: NRC (N = 1301); N L (N = 510); and P I (N = 522).

Each battery consists of the Otis and 2-hand tests in addition to the "Inventory" which gives a name to the battery. Psychiatric criterion.

AREA	BATTERY	N	"GOODS" FAILING PER-CENT	"BADS" N	FAILING PER-CENT	PER-CENT FAILURES IN TOTAL POPULATION	Of the Number Failing, this Percentage Is "Bad"	Chi Square	r t
I	NRC	63	5.7	55	28.5	9.1	47	103.8	.64
	NL	31	7.1	29	38.2	11.7	48	59.3	.69
	PI	24	5.3	18	26.5	8.1	43	35.6	.78
II	NRC	102	9.2	84	43.5	14.3	45	157.9	.67
	NL	45	10.4	40	52.6	16.7	47	82.8	.72
	PI	37	8.2	27	39.7	12.2	42	55.1	.67
III	NRC	200	18.1	123	63.7	24.8	38	183.9	.68
	NL	87	20.0	61	80.3	29.0	41	113.6	.81
	PI	68	15.0	42	61.8	21.1	38	78.1	.71
IV	NRC	313	28.2	146	75.6	35.2	32	161.7	.68
	NL	142	32.7	65	85.5	40.6	31	74.5	.77
	PI	94	20.7	45	66.2	26.6	32	62.6	.67
V	NRC	390	35.2	157	81.3	42.0	29	139.9	.68
	NL	169	40.0	67	88.2	46.3	28	62.9	.76
	PI	133	29.5	48	70.6	34.7	27	44.4	.60

chi square for all levels of all tests is highly significant at negligible levels of p.

We are interested in determining whether the tests, combined in batteries, are more effective in discriminating "Bads" from "Goods" than are individual tests. In order to arrive at a conclusion in this regard, we have constructed Table V, which repeats the data of Table III and in addition gives data for the test batteries. Tetrachoric correlations have been calculated at cut-offs for the individual tests and the batteries which fall nearly comparable percentages of the respective populations.

TABLE V

Tetrachoric Correlations at Cut-Offs Giving Comparable Failure Rates for the Individual Tests and the Batteries

<u>Percentage Failed</u>	<u>NRC Battery</u>	<u>New London Battery</u>	<u>Personal Inventory Battery</u>	<u>2-hand</u>	<u>Otis</u>	<u>NRC</u>	<u>New London</u>	<u>Personal Inventory</u>
7 - 12	.64	.69	.78	.29	.70	.53	.42	.71
12 - 18	.67	.72	.67	.22	.69	.54	.42	.61
24 - 30	.68	.81	.67	.23	.63	.47	.40	.57
Average at 3 levels	.66	.74	.71	.25	.67	.51	.41	.63

If we regard the average tetrachoric correlations in Table V, we may observe that the New London battery and the Personal Inventory battery correlate to a higher degree with criterion (average tetrachoric correlations of .71 and .74, respectively) than any single test, but that the NRC battery correlation with criterion (.66) is exceeded by that for the Otis test alone (.67).

The New London battery seems to be successful in discriminating "bad" men from "good"; particularly at the highest cut-off, and its average tetrachoric correlation with criterion is reliably higher than the correlation for any single test at better than the .01 level.

The Personal Inventory battery does about as well as the New London battery. The correlation for the Personal Inventory battery is reliably higher than the correlation for the Otis test at the .06 level; it is higher than the correlations for the other single tests at below the .01 level. The Personal Inventory does particularly well at the lowest cut-off.

The NRC battery seems to be least discriminative. Its average correlation (.66) with criterion is not reliably higher than the correlations for the Otis test alone (.67) or the Personal Inventory alone (.63). The correlation for the NRC battery exceeds the correlation for the remaining tests at the .01 level. (Reliabilities of differences between correlations were calculated by the reference given in Table VI.)

These results indicate that the combining of tests into the New London and Personal Inventory batteries does increase discrimination above what may be obtained with any single test.

Data on selection cost

Figure 4 gives a graphic representation of results with the 3 batteries. In all cases the Otis - "inventory" combination was preceded in selection by the 2-hand test at a score of 25 or less. Figure 4 shows that as the cut-off level increases, the percentage of "Bads" decreases. In addition, the percentage of men who fail the test increases. The first table to the right of each bar graph indicates the number of men who must be tested to give a class of 100 men at the various cut-off levels. The second column, to the right, gives the number of "bad" men in each class of 100 men thus obtained. If the manpower pool is large enough to provide 172 men for testing, the number of "bad" men in a class of 100 is reduced from 15 to 5 with the NRC battery. With a pool of 141 men, the number of "bad" men is reduced from 15 to 4 (New London battery). Similar results obtain for the Personal Inventory battery.

RESULTS: SUBMARINE MEN TESTED AFTER PATROLIndividual tests and batteries

In the course of our experiments we have tested 161 men from operating crews of submarines. No man in these experiments was a member of a standardizing group. Test results have been analyzed and validated against officers' ratings made on the forms shown in Appendix F.

The present report considers a single validating score obtained from officers' ratings: score on the final "over-all" item.<sup>3</sup>

Table VI contains an aspect of our treatment of results. Men who have been rated unfavorably, 1 or 2 on the final item (rank of 1 most unfavorable), are considered "Bads". Men rated 3, 4, or 5 are considered "Goods" (rank of 5 most favorable). The table presents results for the NRC battery and the New London battery. Data for the NRC battery were obtained on 158 men (126 "Goods", 32 "Bads"). Results for the New London battery were obtained on 92 men (80 "Goods", 12 "Bads"). The final form of the New London questionnaire was used on 32 of the 92 men. Sixty men received the original, extended form which contained more items than the final form. The Personal Inventory was not available for use during most of the experiments of this section, hence no data on this test are presented.

---

<sup>3</sup>This procedure seems justified on the basis of results obtained with other possible validating measures. The correlation (uncorrected) of average score on the first twelve items with score on the final item is .92. Since our primary emphasis is upon temperamental factors, we have also correlated the average score for items which are, from an a priori point of view, not obviously related to intelligence. Average ratings for these items (1, 3, 4, 7, 8, 9, 11 and 12) give a correlation (uncorrected) of .90 with final item. Correlations between each individual item and the final item are listed in Table G1 of Appendix G.

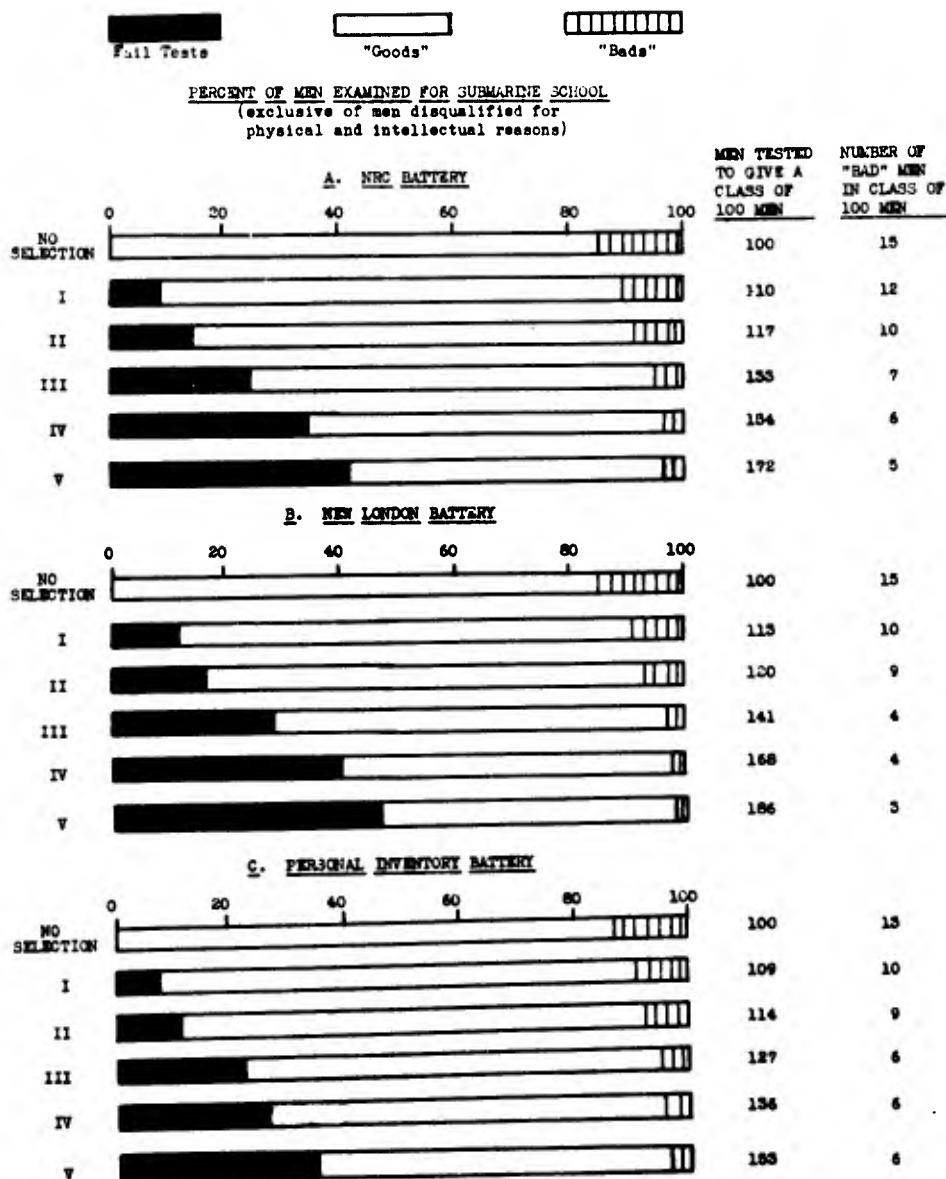


FIGURE 4. Results obtained with the proposed system of combination test score cutoffs. The total length of each bar represents 100 percent of the men tested or examined for Submarine School, exclusive of men disqualified for physical and intellectual deficiencies.

For each "inventory" battery, as the number of men failed by our tests increases, the number of "Bad" men left in the entering Submarine School population decreases.

The number of men who must be tested in order to yield a class of 100 men is listed to the right of the bar graphs, together with the number of "Bad" men remaining in each class after screening at the different "inventory"-Otis combination scores. These figures are given for each of the cutoff areas shown in Figure 3. The number of "Bad" men remaining in each class of 100 men for each cutoff area is compared with the number of "Bad" men in such a class when no screening procedures are used. As the pool of men available for testing increases, the percentage of "Bad" men admitted drops.

TABLE VI

BATTERY RESULTS. OFFICERS' JUDGMENTS. Over-all Item as Validation.

N = 158 for NRC battery. N = 92 for New London battery.

Area	Battery	"GOODS"	Failing	"BADs"	Failing	Percent of Total Pop. Failing	Chi Square	r t
		N	Percent Failing	N	Percent Failing			
I	NRC	11	8.7	9	28.1	12.7	8.4*	.51
	NL	11	13.8	8	66.7	20.7	17.6	.77
II	NRC	16	12.7	11	34.4	17.1	8.3	.47
	NL	16	20.0	8	66.7	26.1	12.0	.68
III	NRC	27	21.4	17	53.1	27.8	12.8	.51
	NL	24	30.0	9	75.0	35.9	9.2	.65
IV	NRC	46	36.5	21	65.6	42.4	8.8	.44
	NL	28	35.0	9	75.0	40.2	7.0	.59
V	NRC	57	45.2	22	68.8	50.0	5.6	.37
	NL	36	45.0	9	75.0	48.9	3.7	.47

\*

Chi square for  $p = .01$  is 6.6 (one degree of freedom)

The tests have been combined according to the method described in the preceding section (Figure 3). Results for the different cut-off areas are contained in Table VI where the values of chi square and the tetrachoric correlations indicate that both batteries discriminate "good" men from "bad" men to a significant degree, but that discrimination by the New London items is more effective.

The question arises as to whether the batteries discriminate "good" men from "bad" men better than does any single test alone. Table VII gives product-moment correlations (corrected for grouping) with officers' over-all rating for the 4 individual tests. These correlations may be compared with the average tetrachoric correlations and biserial correlations for the NRC and New London batteries which are given in the same table. There can be little doubt that the New London battery discriminates "good" men from "bad" better than does any single test alone. The NRC battery is less reliably better than the single tests. It is not reliably better than the New London test alone.

TABLE VII

Correlation of the Individual Tests and the Batteries with Officers' Over-all Ratings. For the batteries, correlations are averages of the tetrachoric correlations obtained at the cut-offs of Table VI. Biserial correlations with officers' judgments are calculated at cut-off II of Table VI. Product-moment correlations, multiplied by 1.06 to correct for coarse grouping, are presented for individual tests.

<u>N</u>	<u>Test</u>	<u>Correlation with Officers' Rating</u>
159	2-hand	.06
160	Otis	-.30
161	NRC	-.22
92	New London	-.34
158	NRC Battery*	.46 (rt) .36 (rbis)
92	New London Battery**	.63 (rt) .65 (rbis)

\*Reliabilities of differences between  $r_t$  and  $r$  for Otis, 2-hand and NRC at less than .05 level.

Reliabilities of differences between  $r_{bis}$  and  $r$  for 2-hand and NRC at less than .02 level.

\*\*Reliably higher than any single test at less than .01 level.

Reliabilities calculated by formula (110), page 188, in Peters, C. C., and van Voorhis, W. R., Statistical Procedures and Their Mathematical Bases. 1940, New York, McGraw-Hill Co. Standard deviations entering into biserial calculations corrected by Sheppard's correction.

#### Data on selection cost

Data for the batteries are given in Figure 5 which shows, for the criterion of over-all item, the number of men failing at each cut-off area, the number who pass at each cut-off area, and the number of "bad" men who are not rejected.

Figure 5 would apply to such a situation as the following: An officer has an opportunity to select 100 experienced submarine men from a large group. He has no information concerning the men. It would be possible on the basis of our batteries to obtain the results indicated in the columns to the right of the bar graphs. The first column indicates the number of men to be tested in order to provide a crew of 100, and the second column indicates the number of "bad" men (ratings 1 or 2) who would remain in a crew of 100 after screening procedures at the 5 different cut-offs were carried out. It is interesting to observe that a pool of 139 men would result in a crew of 100 men for the NRC battery which includes 13 "bad" men, as contrasted with 20 who would exist in the crew if no selection procedures were adopted. Results for the New London battery are more striking. A pool of 126 men would lead to a crew of 100 men containing 5 "bad" men after testing, as contrasted with a crew containing 13 "bad" men when no tests are applied.

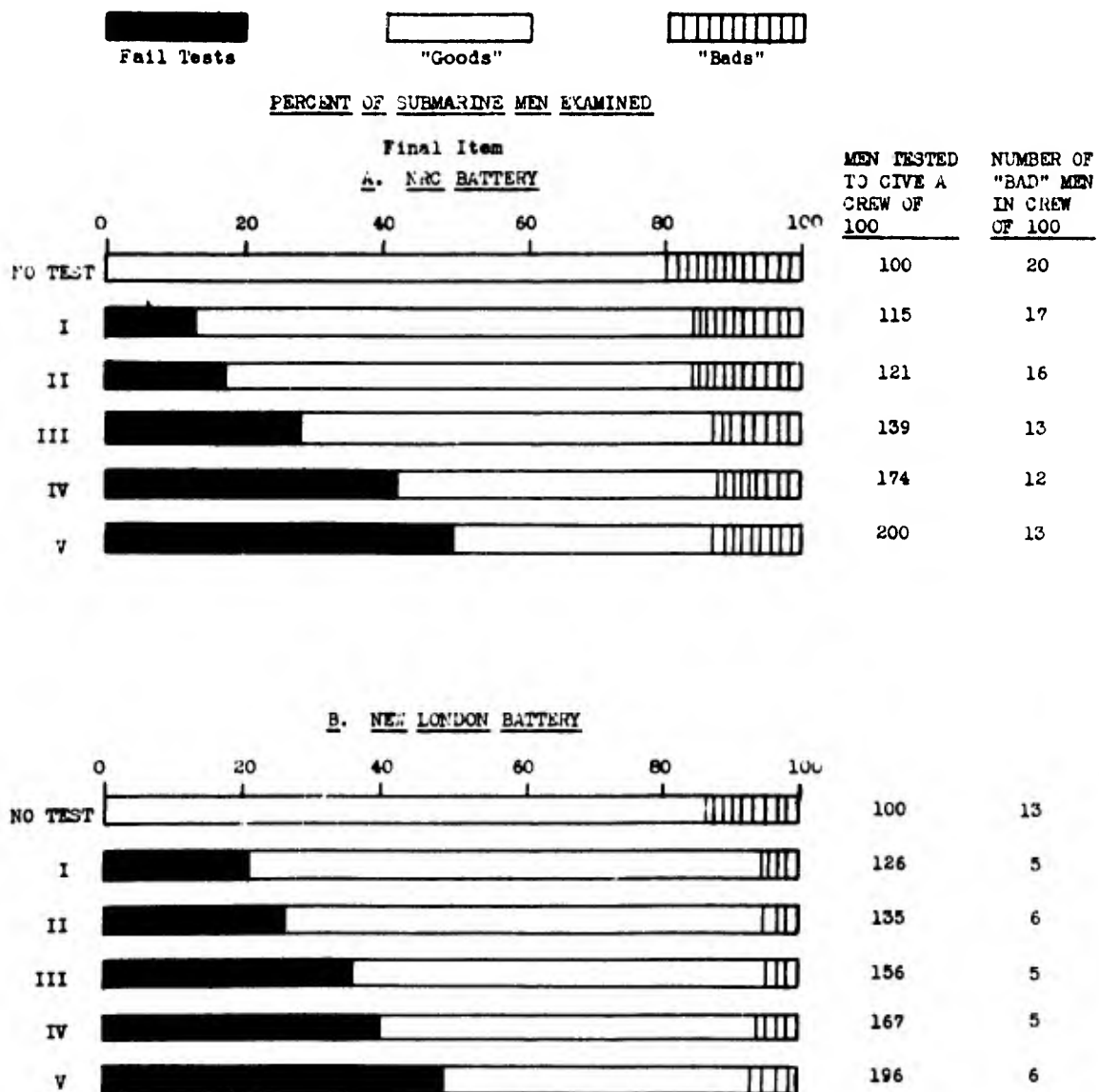


FIGURE 5. Analysis of results for the NRC and New London batteries against final item validation.

RESULTS: TANK PERFORMANCE

Failure in Tank training performance involves such manifestations as:

- (1) Letting go the line and rising rapidly to the surface.
- (2) Spitting out the mouthpiece of the lung.
- (3) Coming up too fast and passing others on the line in an effort to reach the surface rapidly.
- (4) Complaining of pain from pressure, when medical examination shows no indication of any abnormality in the ear.
- (5) Refusing to leave an escape chamber; pleading with the officer in charge to be allowed to wait another day, etc.
- (6) Doubled-up posture in ascent with shaking and shivering.
- (7) Signs of fear during escape training, as shown by shaking, inability to carry out instructions, etc.

Symptoms of the above nature were shown by 42 men in our representative group. The representative group consists of those men, tested by us at the time of the medical examination, who were not disqualified in that examination and thus were allowed to take Tank training.

Although we have some data on the relation of Tank failure to scores on the Personal Inventory battery and the New London battery, the results are so few that our present discussion is restricted to a consideration of the NRC battery.

Experiments on Tank failure are continuing. It is our belief that we shall not obtain adequate data for analysis until results for many more men are available. In this report we shall restrict ourselves to a consideration of two populations in our representative group: (a) a group of 416 men, including 22 Tank "cases" (failures), who contributed to the standardizing of the NRC Inventory (Appendix B); and (b) a group of 1183 men, including 20 tank cases, who were not used for standardizing purposes.

Tables VIII and IX give chi square results for the representative Tank cases. Table VIII shows that, at cut-off areas IV and V (method of Figure 3), a higher percentage of Tank failures are "predicted" in the standardizing group by our tests than could be expected by chance. Chi square values (13.3 and 13.4) at these high cut-offs are reliable at less than the .01 level.

The non-standardizing group does not give comparable results (Table IX). In Table IX the only chi square value which approaches reliability is the value at cut-off area III. The chi square value at this level is reliable at about the .12 level. Chi square values have not been calculated for Areas I and II in either table.

In addition to the Tank cases included in our representative population we have tested 29 men who were referred to us after they had shown abnormal Tank performance. These men had not received our tests prior to Tank training. Comparison of the bottom and top rows of Table X shows that these men (referred Tank cases) evidenced a much higher failure rate on our tests

TABLE VIII

NRC BATTERY RESULTS. REPRESENTATIVE TANK CASES.  
STANDARDIZING CASES.

"Goods", 394; "Bads" (=Tank Failures), 22.

Area	"Goods"		"Bads"		Percent of Total Population Failing	Chi Square
	N	Percent Failures	N	Percent Failures		
III	88	22.3	9	40.9	23.3	4.1
IV	122	31.0	15	68.2	32.9	13.3
V	150	38.1	17	77.3	40.1	13.4

$\chi^2 = 6.6$  for  $p = .01$  (one degree of freedom)

TABLE IX

NRC BATTERY RESULTS. REPRESENTATIVE TANK CASES.  
NON-STANDARDIZING CASES

"Goods", 1163; "Bads" (=Tank Failures), 20.

Area	"Goods"		"Bads"		Percent of Total Population Failing	Chi square
	N	Percent Failures	N	Percent Failures		
III	247	21.2	8	40.0	21.6	4.1
IV	363	31.2	9	45.0	31.4	1.7
V	432	37.1	11	55.0	37.4	2.7

TABLE X

PERCENTAGE OF TANK CASES FAILING  
EACH SCREENING CUT-OFF

	I	II	III	IV	V
Expected by chance*	7.2	11.9	21.6	31.4	37.4
Representa- tive Tank Cases**	4.6 20.0	10.2 25.0	40.9 40.0	68.2 45.0	77.3 55.0
Referred Tank Cases	31.0	34.6	51.8	62.1	62.1

\*Non-standardizing group.

\*\*Upper row, standardizing group; lower row, non-standardizing group.

at all cut-off areas than would be expected by chance (chance expectancy figures based on the non-standardizing group). The middle row in Table X repeats results of Tables VIII and IX. The upper set of figures gives the percentage of failures at the various cut-offs for the Tank cases in the standardizing group, and the lower row, percentages for the non-standardizing group. It may be observed that the failure percentage at cut-offs III, IV, and V for the representative Tank cases and the referred Tank cases seem to be higher than might be expected by chance. Further study of this problem is required.

## DISCUSSION

### Evaluation

Early in our account we stated that the present report must be considered a progress report rather than a final report. A survey of our experimental results will indicate that many things remain to be done.

We have been hampered in certain phases of our work, particularly work on Tank failure and on officers' ratings of men from operating crews, because the only "inventory" available to us from the beginning was the NRC. Thus, a considerable amount of our time and energies went into the construction of the New London Questionnaire. This development involved standardization and revision, with a consequent decrease in the number of men available for strictly experimental procedures. In addition, it was only in the last stages of our work that the Personal Inventory became available. Because of a paucity of "inventories" in the initial phases of the work our results on submarine men from operating crews and on Tank failure are not as complete as are those on psychiatric criterion. Considerable work may be done in the future on the analysis of results obtained with the New London and Personal Inventories.

It is hoped that the work may continue to a satisfactory conclusion. Certainly we should hope to be able to make an analysis of officers' ratings of testee performance after the testees have joined the Fleet. In addition, further data on Tank failure may be studied to advantage, as well as the performance of operating crews of submarines. Recently the New London Questionnaire has been given at Newport and on a battleship, and it is hoped that valuable information may be obtained from these additional sources. In the event that the recommendations made in this report are acceptable to officers at the Submarine Base, it will probably become important to analyze preliminary results which are obtained with the selection procedures adopted.

The results of the present experiment indicate that our batteries do, in fact, predict two criteria of "psychological unfitness" in significant and usable degree (psychiatric judgment and officers' ratings). The extent to which the tests may be used and the particular method of using the tests will, of course, depend on the particular personnel requirements of a given service branch. Since the experiments were performed at the Submarine Base, we may use that activity as an example of possible applications.

### Selection methods

1. Our results indicate that if a certain amount of "selection loss" be allowable, the number of "bad" men admitted to the Submarine School may be drastically reduced. For example, Figure 4 shows that when the pool of available men is increased to about 150, the number of "bad" men in a class of 100 men is reduced from 15 to 4 or 5. Results obtained on men from

operating crews of submarines support this general type of result.

Although the procedure outlined above would certainly result in a decrease in the number of "bad" men in the Submarine School population, it is well to recognize that the selection loss entailed might be so great as to make the procedure undesirable. Specifically, it would be necessary to supply 150 men in order to obtain a class of 100.

2. Another procedure has certain advantages over the one suggested above. It has the advantage of not requiring a greater pool of men than is required at the present time. The effectiveness of the program lies in the fact that it would relieve New London psychiatrists of about two-thirds of their present selection duties and thus make them available for other pressing activities.

The recommended procedure follows.

All candidates entering the Submarine School might take, at the time of their medical examinations, one of the batteries described in the present report.

It is proposed that all men who fail the accepted test battery, at an appropriate cut-off area, be sent to the psychiatrist for interview. Thus, on the basis of his test results, computed in accordance with the procedure of Figure 3, a candidate may be assigned to or excused from the psychiatric interview. We suggest the adoption of a cut-off which allows a test failure rate of about 30 percent.

As an example of the application of the procedure, let us consider what might happen at a cut-off area which results in a failure rate lying near the desired figure of 30 percent. The New London battery at a cut-off of III fulfills these conditions. (It is probable that the Personal Inventory battery will require a readjustment of cut-off in order to include more failures).

Table IV shows that cut-off III for the New London battery results in a test failure rate of 29 percent of the total population. Thus, according to our proposal, 29 percent of the Submarine School candidates would be sent to the psychiatrist for interview. Included within the 29 percent are 80 percent of the "Bads". This figure indicates that not all of the "bad" men are "caught", and it may be assumed that 20 percent pass on to School activities. However, the figure of 80 percent identification seems to be satisfactory when measured against the standard represented by the careful analysis of interview "successes" and "misses" given by Wittson, Harris and Hunt.<sup>4</sup>

"Good" men, too, would be present in the group referred for interview. However, the latter would not be penalized, since they presumably would be passed.

It is worth pointing out that the procedure outlined above may have an additional advantage. Some "bad" men, who might ordinarily pass the

---

4

Wittson, C.L., Harris, H.I. and Hunt, W.A. An evaluation of the brief psychiatric interview. J. Psychol., 1943, 16, 107-114.

interview, may be identified because of the additional information provided by our tests. It is possible that the advantages in this regard might counterbalance the disadvantages inherent in the expected "misses".

A final step in selection would come when the senior medical officer correlates all of the data on the man's record. Since the record would be extended by our data, we believe that the test procedure suggested would not only provide an efficient aid to the psychiatrist, but, under certain circumstances, could provide enhanced effectiveness through additional information.

A1

APPENDIX A

The Shock-Fusion Test

SUMMARY

A method is described for investigating the effects of "emotional stress" on reaction time for binocular fusion.

The apparatus consists of a visual system, an electric circuit for registering reaction time, and circuit for applying shock to the subject.

The visual system is enclosed in a portable box. The subject looks into the darkened interior of the box and is presented with two spots of green light. Due to the employment of Polaroid, one of the spots stimulates the right eye, and the other, the left eye. The subject's problem is to fuse the two spots into a single spot as rapidly as possible after the two-spot stimuli flash out of the darkness. Reaction time is recorded by means of an electric timer for one condition of fusion, i.e., when the spots are physically close together in space. "Blank" trials involve the presentation of greater separation of spots and hence lower degrees of convergence. A depolarizing screen allows for presentation of both spots to each eye. In this way the experimenter can make sure that the subject responds on the basis of fusion and not to the appearance of light alone.

Certain trials are control trials and involve no shock or threat of shock. Shock is presented in some trials, and in others, the apprehension trials, the threat of shock is present (due to the presence of electrodes and the temporal juxtaposition of shock trials).

The most promising measure of performance obtained in this situation is the Q score, the interquartile range of reaction times throughout the series.

Seventy-five representative Submarine School men give an odd-even trial reliability coefficient for Q of 0.72. This figure is raised to 0.84 on application of the Spearman-Brown formula.

Statistical analyses have been performed to determine the efficiency of Q in differentiating between "good" and "bad" men (N=377).

On the basis of a cut-off score of  $Q = 0.10$ , 24 percent of the "Goods" ("Normals" and "Questionables" by psychiatric judgment) are included in the rejected category (scores higher than 0.10), while 52 percent of the "Bads" ("Psychiatric Bads" and "Tank Failures") are rejected.

VISUAL STIMULATION

The apparatus employed in visual stimulation is housed in a shallow box at the eye level of the subject. It is designed so that each of the subject's eyes is stimulated by one of two small circular spots of light; the subject's task is to fuse these spots when their physical separation is varied. This dichopic stimulation, which is effected by the use of Polaroid film, required different degrees of convergence of the subject's eyes as the separation between lights is varied.

A top view of the visual apparatus is shown in Figure A1.. The subject's head is positioned by means of the hood at the left of the diagram. Immediately in front of his eyes are green glass filters (Corning H. R. Emerald Green, D glass) and squares of Polaroid "H" glass. A transverse partition supports two flashlight units, each of which is viewed by one eye of the subject. Each unit consists of a cylindrical housing within which are mounted a circular diaphragm ( $\frac{1}{4}$ " in diameter), a disc of Polaroid "H" film, a single flashed glass diffusing

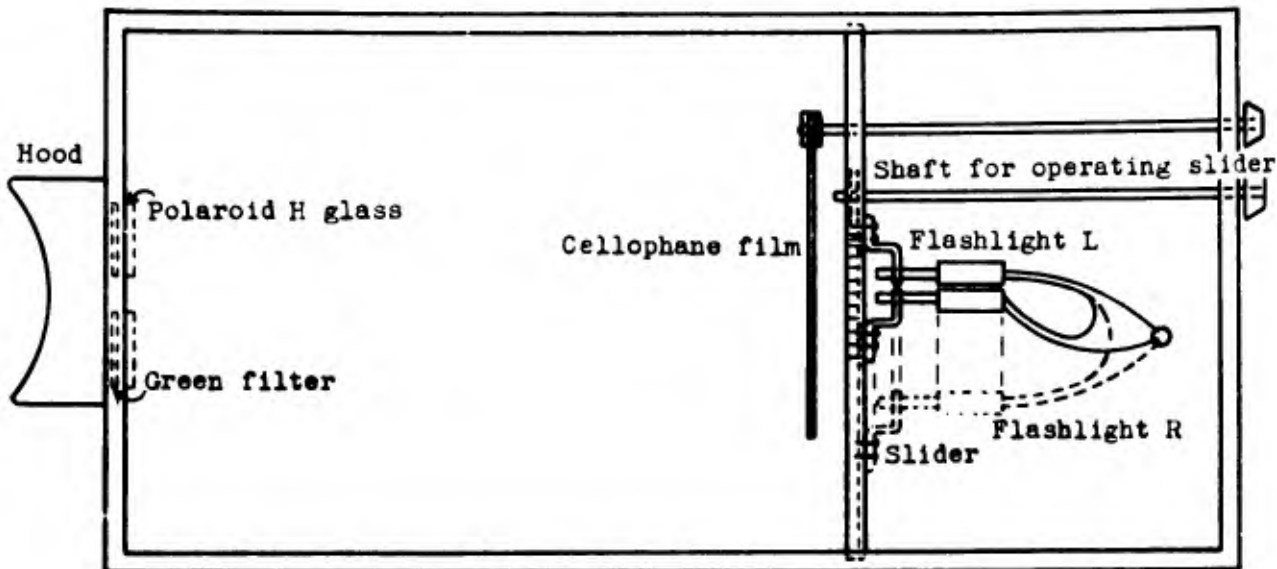


FIGURE A 1

A diagram of the visual apparatus used in the reaction time for fusion experiment.

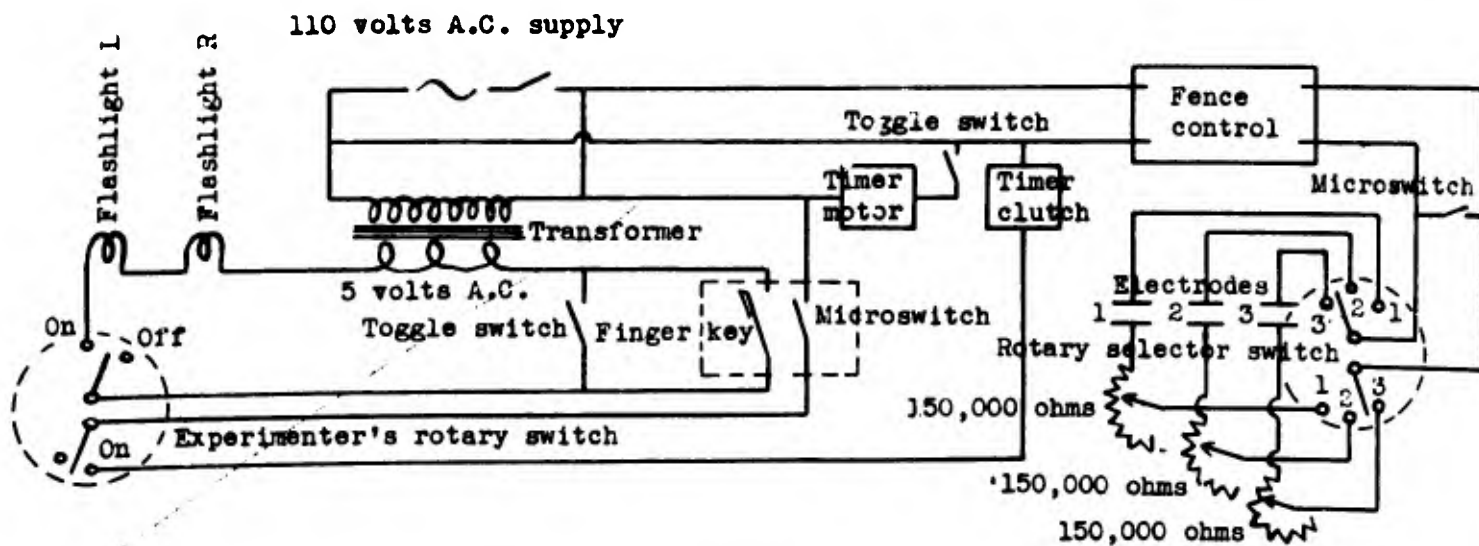


FIGURE A 2

Electrical circuits used in the reaction time for fusion experiment.

screen, and a 3.8 volt flashlight bulb.

The "H" glass in front of each eye is polarized in the same plane as that of the "H" film in the corresponding flashlight housing, and in the opposite plane from that of the film in the other housing. Each eye sees, therefore, but a single circular patch of light; and with the proper degree of convergence the subject is able to fuse the circles viewed by the two eyes.

The flashlight stimulus for the right eye is mounted upon a slider which moves laterally and changes the separation between the patches of light. The separation is varied between the limits of  $\frac{1}{2}$  inch and 3 inches by means of a knob and dial arrangement working through a rack and pinion on the slider. An auxiliary unit, also operated by a knob at the back of the instrument, consists of a sheet of cellophane which may be inserted between the flashlights and the eyes of the subject. The effect of the cellophane is to rotate the polarized light sufficiently so that each eye sees both patches of light. Under these conditions the subject may legitimately report seeing four, three, or two patches of light (depending upon his state of fusion), but never one. This unit has been found to be necessary in order to detect fraud by a subject whose finger reactions are made before fusion has actually occurred.

#### ELECTRIC CIRCUITS

The electrical connections are shown in the diagram of Figure A2. The circuits are supplied by the conventional 110-volt alternating current.

a. The flashlight circuit. The flashlights for the left and right eyes are wired in series and run on alternating current of 5 volts supplied by a transformer. A finger reaction key for the subject and a rotary switch for the experimenter are wired in series so that both must be closed in order for the stimulating lights to be on. In operation the subject prepares for a reaction by depressing the finger key. The experimenter is then able to turn on the lights through the use of his rotary switch. When the subject fuses the lights, he releases the finger key and the lights go off.

b. The timer circuit. Reaction times are measured, in units of 0.01 sec., by a Standard Electric Timer. The interval-registering clutch mechanism is in operation only during the time that the stimulating lights are on. The clutch is activated by an electrical circuit which supplies alternating current to it through one arm of the experimenter's rotary switch and a microswitch which is connected mechanically with the subject's reaction key. Thus the timer registers the interval between the onset of visual stimulation and the subject's signal that he has fused the two lights.

c. The electric shock circuit. A modification of a commercially available electric fence control (Lafayette Radio Corp., N.Y.C., Cat. #K15674) is used for electrical stimulation of the subject. This instrument employs a gas discharge tube and a capacitance to produce a relaxation oscillation which supplies brief monophasic pulses with a peak in the neighborhood of 600 volts. The rate of discharge, 1 to 2 pulses per second in the commercial instrument, has been increased to approximately 10 pulses per second by replacing the standard condenser with one of lower capacity.

Electrical stimulation is delivered to the subject's skin through silver disk electrodes with the aid of a saline electrode paste. The electrodes are held against the skin by means of adjustable rubber straps. Each of the three pairs of electrodes is in series with a variable resistance from 0 to 150,000 ohms for controlling the intensity of stimulation. A rotary selector switch

enables the experimenter to restrict the stimulation to one electrode pair at a time. Stimulation is accomplished by opening a microswitch, which is normally closed, to short-circuit the output of the fence control unit.

d. The resistance bridge. A simple Wheatstone bridge circuit, independent of the other electrical circuits, is used for the measurement of skin resistance. A microammeter in this circuit may be used as a null instrument so that the approximate skin resistance may be read from the position of the dial on the variable resistor. The meter also provides a rough indication of the magnitude of any galvanic skin responses.

#### SUPPLEMENTARY APPARATUS

Inter-trial interval, warning signal, and shock duration are timed by means of a stop watch.

#### PROCEDURE

Figure A3 gives a photograph of the experimental situation.

As soon as the subject is seated in the chair in front of the instrument, the electrodes of the instrument for measurement of the galvanic skin resistance are put on the palm and back of the left hand, the subject being assured that these are completely inactive and that he can "trust" them.

The following procedure is then followed: The subject is instructed to place his head in the hood, and to tell what he sees. The light is then turned on with the spots at "near" separation. If the subject fails to mention how many there are, he is asked, "How many spots do you see?" Extreme "far" separation is then introduced, yielding invariably a response "Two". Now intermediate separations are introduced, until the experimenter has succeeded in demonstrating to the subject the phenomenon of fusion (the movement together of two spots until they become one) and, if possible, until the experimenter has determined a separation for which fusion reaction time is relatively great. When this separation has been determined, the following instructions are given: "This is an experiment in reaction time. You will observe at your right hand a telegraph key. Be careful not to handle the metal parts of this key. Our procedure will be as follows: I will give you a 'ready' signal, at which you are to press down on the key, and place your head in the hood. In four seconds the light will come on. As soon as you see one spot of light, release the key as fast as you can. As long as you see two or more lights, keep the key down. As soon as you have responded, or as soon as the light goes out, remove your head from the hood, and look over to the left hand corner of the room. In twenty seconds the next trial will begin. If you make an error, that is, if you release the key while you still see two spots, tell me so, and it will be all right. Remember that you must be both fast and accurate. As soon as you see one spot of light release the key as fast as you can. Are there any questions?"

The practice series is then begun. Twenty trials, 10 "ins" and 10 "outs", are given. On some "out" trials the depolarizing screen is placed in position and the subject's accuracy checked. If the previously determined "out" separation is proved unsatisfactory in these trials, a new separation is used. <sup>Al</sup>

<sup>Al</sup> Either of these procedures may be used on any "out" trial during the experiment. The second has proved necessary for the majority of subjects, all of whom show great variability on the "out" separation. The first procedure is necessary for those subjects who seem to be responding consistently too fast, i.e., fallaciously on all "out" trials. An objective check on accuracy is necessary in these cases. The original separation often becomes either too difficult or too easy for satisfactory use.

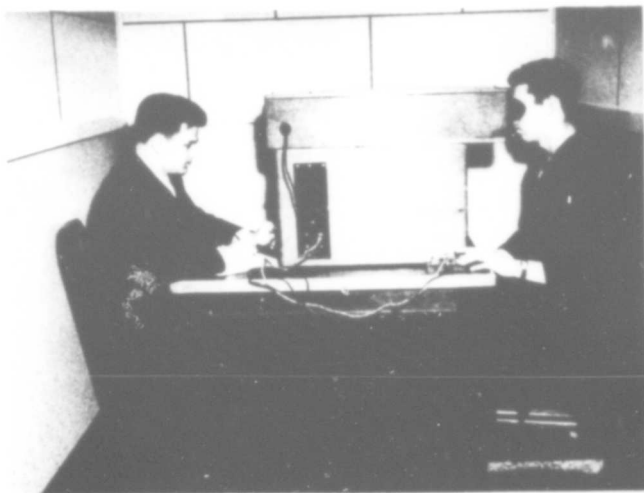


FIGURE A 3. The shock-fusion apparatus.

When these twenty practice trials are completed, the subject is informed that a three minute rest period is to come. At the end of three minutes, the first control series of 5 "ins" and 5 "outs" is run. The GSR electrodes are now removed, and the subject is instructed to remove his shirt, and to seat himself on a chair at the side of the apparatus. Three pairs of electrodes are now fixed to the subject, one on the calf of the right leg, one on the left shoulder, and one on the left forearm. The high member of the latter pair is used in the determination of shock intensity to be applied. The intensity of shock used is that necessary to produce a vigorous contraction of M. abductor pollicis brevis and M. flexor pollicis brevis, elicited by cathodal stimulation of the mons of the thumb. When this intensity of stimulation has been determined, the subject is reseated at his place in front of the apparatus and the GSR electrodes are replaced. The experimenter now equates the intensity of shock through the shoulder and leg electrodes to that of the forearm on the basis of the magnitude of the GSR, subject's overt response, and subject's verbal report. Upon the completion of this procedure, the following instructions are given: "From this point on, you will be shocked on certain trials. The shock will come after I have given the 'ready' signal, but before the light comes on. Your job is to do exactly as you have been doing. Remember, the shock will come on some trials and at some place after the 'ready' signal and before the light comes on. Otherwise the procedure is the same".

The experimental series then begins. Ten trials during which no shock is given are followed by 12 trials on half of which shock occurs in random order. Eleven trials, all of which are accompanied by shock are then presented, followed by 12 random shock and finally 10 no-shock trials. There immediately follows a three minute rest during which the electrodes are removed. The subject is further assured verbally that there will be no shocks. After this rest, the final series of 10 control trials is given. Throughout the series "ins" and "outs" are given in random order, and in equal number.

The warning interval employed is four seconds, the inter-trial interval, 15 seconds. On the trials in which shock is delivered, the shock lasts approximately through the third second of the warning interval. Whenever the subject fails to respond within 5 seconds of the onset of stimulation, the stimulus is withdrawn, and on the next trial a new and smaller inter-spot distance is presented.

#### SOME RESULTS

A number of calculations has led us to accept the Q value for all "in" trials as a fruitful measure of performance. The Q value is simply the semi-interquartile range for all the "in" readings and hence is a measure of variability of performance. It is possible to compute values for various fractional parts of the series ("apprehension" trials, shock trials, control trials) but intercorrelations indicate that the over-all variability measure, Q, is best. Q has the advantage of disregarding occasional, a typical, long reaction times and it is generally more convenient than the standard deviation. However, the correlation between the standard deviation (with Chauvenet's criterion applied) and Q is over 0.90 (48 soldiers examined during October and November, 1941, at Fort Monroe, Virginia).

Seventy-five representative Submarine School men give an odd-even trial reliability for Q of 0.72. This figure is raised to 0.84 on application of the Spearman-Brown formula.

Table A1 indicates something of the validity of the shock-fusion test.

The "Good" and "Bad" groups are composed as described in the Summary of this Appendix. A cut-off Q score of 0.10 is used. Data are for men tested up to about November 15, 1942. (Results are not available for men after this date because of our unsuccessful attempts to simplify the test procedure.)

Table A1

## Tabulation of shock-fusion data

	"Goods"	"Bads"	Totals
0-0.09	232	34	266
0.10 --	74	37	111
Totals	306	71	377

The table indicates that a cut-off of 0.10 rejects 52 percent of the "Bads" and 24 percent of the "Goods". Chi square is 21.7, a value which shows good differentiating properties for the test. Discrimination is continued at a satisfactory level at higher cut-offs (Chi square = 22.3 at 0.12; 23.0 at 0.14).


Because of the excessive time required for its administration (45 minutes), the test was discontinued in our program after December, 1942.

B1



APPENDIX B

Item Analysis of the NRC Neurotic Inventory: Submarine School Population



Using a large population, which became available in the course of our experiments, we have made an item analysis of responses for the NRC Neurotic Inventory (NRC Form R-2).

The population consisted of 728 Submarine School men. Of these, 578 were rated as either "Normal" or "Questionable" by the psychiatrist; 127 were rated "psychiatric Bads" and 23, "Tank failures". "Psychiatric Bads" and "Tank failures" were combined to make up the validating group ("Bads"). The "Normals" and "Questionables" were combined to form the "Goods".

The analysis has been made by contrasting the performance of "Normals" and "Questionables" with the performance of "Bads" and "Tank failures". Frequencies of response given by each class to the various items on the inventory were determined by means of a machine scoring method, and reliabilities of differences were determined by the Fulcher-Zubin Nomograph (J. Appl. Psychol., 1942, 26, 511-522).

Table B1 summarizes the composition of groups and gives a description of results and method. Table B2 indicates the selective items. Each selective item answered contributes a point to the total sum.

Percentile norms have recently been determined on a new population of 1929 men. Table B3 gives reference percentile values.

Table B1

Composition of Standard and Validating Groups  
for Item Analysis,  
together with Description of Results and Method

<u>Standard Population</u>	<u>Validating Cases</u>	<u>Character of Finding</u>	<u>Method</u>
578 Submarine men ("Normals" and "Questionables")	150 Submarine men ("Buds" and "Tank failures")	41 items	Fulcher-Zubin Nemograph
		8 items	
		P<0.01	
		9 items	
		P=0.01-.02	
24 items			
		P=0.02-.05	

Table B2  
Selective Items Based on Item Analysis

4 T	6 F	13 F	15 T	16 T	18 T	19 F	22 F
23 T	29 F	31 F	33 T	36 T	37 F	41 T	49 F
51 T	52 T	55 F	57 F	59 F	60 F	61 T	66 T
68 F	72 T	82 F	101 T	109 T	112 F	113 F	116 T
118 T	123 F	130 F	135 F	139 T	142 T	146 T	147 F
149 T							

Table B3

Reference Percentiles for Different Scores  
(New Population)

<u>Score</u>	<u>Percentile</u>	<u>Score</u>	<u>Percentile</u>	<u>Score</u>	<u>Percentile</u>
5	1	12	50	17	89
7	5	13	60	19	95
8	10	14	69	20	97
9	18	15	78	21	96
10	27	16	84	22	99

C1

APPENDIX C

The New London-NDRC Confidential Questionnaire


NEW LONDON - N. D. R. C.

CONFIDENTIAL QUESTIONNAIRE

\*\*\*\*\*

This booklet is not a test. It is a group of questions concerning the facts about you and your life. Each question is asked so that you should be able to answer it. Sometimes none of the answers will be the exact one you would choose; in such a case, choose the answer that most nearly describes you.

Choose one answer, and only one answer, to each and every one of the questions. Do not mark this booklet in any way. Indicate your answers on the answer sheet.



████████████████████

	country	1
	village	2
1. I was raised in:	town	3
	city	4
<hr/>		
	is living	5
	died within the last year	6
2. My own mother:	died from 1 to 5 years ago	7
	died more than five years ago	8
<hr/>		
	is not living	9
	is living with my father	10
3. My own mother:	is separated from my father	11
	is divorced from my father	12
	is divorced and remarried	13
	is single after my father's death	14
	is remarried after my father's death	15
<hr/>		
	is living	16
4. My own mother:	died of illness	17
	died of accident	18
	died of old age	19
<hr/>		
	is living	20
5. My own father:	died of illness	21
	died of accident	22
	died of old age	23
<hr/>		
	is not living	24
	is living with my mother	25
	is separated from my mother	26
6. My own father:	is divorced from my mother	27
	is divorced and remarried	28
	is single after my mother's death	29
	is remarried after my mother's death	30

████████████████████

	living together	31
	separated more than 5 years	32
	separated less than 5 years	33
7. My own parents are:	divorced more than 5 years	34
	divorced less than 5 years	35
	not both living	36
	neither living	37
	the only child in the family	38
8. I was: (Do not count any brothers or sisters who died before one year of age.)	the eldest child	39
	the youngest child	40
	neither the oldest nor youngest	41
	well-to-do	42
9. I was brought up in circum- stances which were:	comfortably off	43
	poor	44
	very poor	45
	own parents	46
	own mother alone	47
	own father alone	48
	one parent and one stepparent	49
10. I was brought up by:	two stepparents	50
	grandparents	51
	aunts and uncles	52
	friends	53
	homes	54
	elder brothers and sisters	55
	I wanted to help win the war	56
11. I chose this service (Army, Navy or Marines) because:	I preferred this service to the others	57
	I wanted to get away from home	58
	I had trouble in civilian life	59
	I always wanted to	60

- college 61
- part college 62
- high school 63
12. I finished:
- part of high school 64
- only the 9th grade 65
- 
- grammar school 66
- was expelled or suspended 67
- didn't like it 68
- was getting poor grades 69
13. I left school the last time because I:
- needed the job to help at home 70
- graduated 71
- joined the Navy 72
- 
- would have been expelled 73
14. I was left back for one reason or another a half year or more in grammar school or high school:
- yes 74
- 
- no 75
- good grades 76
15. I received:
- average grades 77
- 
- poor grades 78
- close friends 79
- one or two close friends 80
16. When I was in school, I had:
- a small number of close friends 81
- a lot of close friends 82
- 
- many acquaintances 83
- arguments 84
17. I had the following trouble with my fellow students:
- fights 85
- 
- none 86
- frequently 87
18. I stayed away from school for reasons other than illness or work:
- sometimes 88
- 
- never 89

████████████████████

C6	and have always been single	91
	going steady	92
	engaged	93
19. I am:	married	94
	divorced	95
	separated	96
	divorced and remarried	97
	<u>not engaged, although I was at one time</u>	<u>98</u>
	to remain married	99
	to get a separation or divorce	100
	never to get married	101
20. I expect:	to get married after I leave the Navy	102
	to get married after the war	103
	<u>undecided whether to marry or not</u>	<u>104</u>
		105
	more than once a year	106
	about once a year	107
21. Before joining the Service, I changed jobs:	about every two years	108
	about every four years	109
	never	110
	<u>have never had a steady job</u>	<u>111</u>
	I got a better job	112
	I didn't get along with my boss	113
	I was laid off	114
22. I changed jobs because:	I was discharged	115
	I joined the Navy	116
	<u>I didn't have to work</u>	<u>117</u>
		118
		119
		120

████████████████████

		I was home	121
23.	Of the following things, I would like most to have there was a good house for mother and father this one happen to me:	I could have my girl back	122
		I wish I could have a successful life	123
		I could have lots of money	124
<hr/>			
		movies	125
		reading	126
		cars	127
24.	Of the following hobbies, my favorite would be:	letter-writing	128
		listening to the radio	129
<hr/>			
		shot-put, track, boxing, wrestling	130
		tennis, horseshoes, badminton, etc.	131
25.	I like best to indulge in such sports as:	baseball, basketball, football	132
		hunting, fishing, hiking	133
<hr/>			
			134
26.	I read Time, Readers' Digest, Life, Newsweek, etc.:	yes	135
		no	136
<hr/>			
		Colliers, Liberty, Sat. Eve. Post, American	137
		Picture magazines	138
27.	My favorite magazines, besides the ones already named are:	Love or romance (True Story, Confessions, etc.)	139
		Detective, Sports, Western	140
		Mechanical or trade (Pop. Mechanics, etc.)	141
		Comics (Superman, Dick Tracy, etc.)	142
<hr/>			
		ever since I could drive	143
		since I was sixteen	144
		since I was eighteen	145
28.	There has usually been a car for me to drive:	since I was twenty-one	146
		never	147
		don't drive	148
<hr/>			
			149
			150

29. I have been called down by the police for a traffic offense:

once 151  
twice 152  
3 times 153  
more than 3 times 154  
never 155

---

30. I have been picked up by the police for other offenses:

once 156  
twice 157  
3 times 158  
more than 3 times 159  
never 160

---

31. I have been convicted of these other offenses

once 161  
twice 162  
3 times 163  
more than 3 times 164  
never 165

---

32. I have been picked up by the Shore Patrol or by the Military Police:

once 166  
twice 167  
3 times 168  
more than 3 times 169  
never 170

---

33. I have been on report:

once 171  
twice 172  
3 times 173  
more than 3 times 174  
never 175

---

34. I have had a Summary Court Martial:


once 176  
twice 177  
3 times 178  
more than 3 times 179  
never 180

	sleeping in	181
	neglect of duty	182
	AOL or AWOL	183
	drinking	184
35. I have usually been in trouble because of:	insubordination	185
	fighting	186
	have never been in trouble	187
<hr/>		
	invest it	188
36. If I had a million dollars that I could spend any way I wanted, the first thing I would do with my money would be: (Choose one from this group)	buy a business	189
	give it to my mother	190
	give it to charity	191
	buy a car	192
<hr/>		
	I would receive more money	193
37. I would like an increase in rate or a promotion in rank because:	I would have more authority	194
	I would have fewer persons to work for	195
<hr/>		
	yes	196
38. I often lose out in something by not making up my mind quickly enough:	no	197
	don't know	198
<hr/>		
	yes	199
39. I usually like to have others plan my work for me:	no	200
	don't know	201
<hr/>		
	very frequently	202
	frequently	203
40. I run into people so unreasonable that I find it hard to talk to them:	occasionally	204
	seldom	205
	never	206
<hr/>		
	immediately	207
	by myself after some thought	208
41. I usually make decisions:	after getting advice from others	209
	with great difficulty	210

42. If I were in command of a detail, I would expect my detail to carry out every order without comment:	yes	211
	no	212
<hr/>		
43. I have been fairly seriously injured in an accident or battle:	never	213
	once	214
	twice	215
	3 times or more	216
<hr/>		
44. I have been seriously injured: (broken bones, concussion, severe cuts or bruises)	once	217
	twice	218
	3 times	219
	more than 3 times	220
	never	221
<hr/>		
45. I usually eat:	heartily	222
	a good meal	223
	moderately	224
	very little	225
<hr/>		
46. I have upset stomach:	frequently	226
	occasionally	227
	rarely	228
<hr/>		
47. My bowels move:	never	229
	once a day	230
48. I am subject to constipation:	twice a day	231
	irregularly	232
	frequently	233
<hr/>		
49. I am subject to loose bowels:	occasionally	234
	rarely	235
	never	236
<hr/>		
49. I am subject to loose bowels:	frequently	237
	occasionally	238
	rarely	239
<hr/>		
49. I am subject to loose bowels:	never	240

	once a week	241
	once every 2 weeks	242
50. I have headaches:	once a month	243
	once a year	244
	never	245
<hr/>		
51. I have had scarlet fever:	yes	246
	no	247
<hr/>		
52. I have had St. Vitus Dance:	yes	248
	no	249
<hr/>		
53. I have had whooping cough:	yes	250
	no	251
<hr/>		
54. I have had measles:	yes	252
	no	253
<hr/>		
55. I have had infantile paralysis:.	yes	254
	no	255
<hr/>		
56. I have had malaria:	yes	256
	no	257
<hr/>		
57. I have had typhoid fever:	yes	258
	no	259
<hr/>		
58. I have or have had tuberculosis:	yes	260
	no	261
<hr/>		
59. I have or have had epilepsy (convulsions or fits):	yes	262
	no	263
<hr/>		
60. I have had pneumonia:	yes	264
	no	265
<hr/>		
61. I have or have had syphilis:	yes	266
	no	267
<hr/>		
62. I have or have had gonorrhoea:	yes	268
	no	269

	I have never had intercourse	271
	I have never exposed myself with a possibly diseased girl	272
63.	I have never had a venereal disease because:	
	I was careful to use a rubber	273
	I used prophylaxis immediately after exposure	274
	I was lucky	275
	<hr/>	
	I have had a venereal disease	276
	to excess sometimes	277
64.	I drink:	
	never drink to excess	278
	<hr/>	
	don't drink	279
65.	After I have had three or four drinks:	
	I would just as soon drive as not	280
	I would rather not drive	281
	<hr/>	
	I don't drink or don't drive	282
	never	283
66.	I have bitten my nails:	
	for the last few years	284
	<hr/>	
	in the past but not now	285
	<hr/>	
	everything set before me	286
67.	I eat:	
	almost all foods	287
	most foods	288
	<hr/>	
	only certain foods	289
	once in two years	290
	once a year	291
68.	I have colds:	
	once in six months	292
	once in three months	293
	<hr/>	
	never	294
	every day	295
69.	I become very nervous:	
	once or twice a week	296
	once or twice a month	297
	<hr/>	
	never	298
		299
		300

		
		in the morning 301
		in the middle of the day 302
70. When I become nervous it is usually:		in the evening 303
		at no particular time 304
		never 305
<hr/>		
		nearly all the time 306
		about once a week 307
71. I have periods when I feel sad:		about once a month 308
		about once a year 309
		never 310
<hr/>		
		a few hours 311
		several days 312
72. These periods of sadness usually last about:		several weeks 313
		several months 314
		never have any 315
<hr/>		
		to lose weight 316
		to lose my appetite 317
		to lose sleep 318
73. These periods of sadness cause me:		to worry 319
		to have nightmares 320
		never have any 321
<hr/>		
		go to bed 322
		take a coke 323
		have a few beers 324
		can't sleep 325
74. When I am tired, I usually: (Choose only one.)		go to movies 326
		get a headache 327
		get nervous 328
		get short tempered 329
<hr/>		



	right handed	331
75. I am	left handed	332
	both	333
	a good driver	334
	a safe driver	335
	a clever driver	336
76. With respect to driving an auto- mobile, I am considered to be:	a daring driver	337
	a skillful driver	338
	a poor driver	339
	don't drive	340
	very well	341
77. I dance:	fairly well	342
	poorly	343
	not at all	344
		345
78. I play some on one or more musical instruments:	yes	346
	no	347
	very well	348
79. I sing:	fairly well	349
	poorly	350
	not at all	351
80. I have repaired broken electrical gear:	yes	352
	no	353
81. I have repaired broken clocks or watches:	yes	354
	no	355
82. I have repaired broken furniture:	yes	356
	no	357
		358
		359
		360

D1

APPENDIX D

Scoring Sheet for the New London-NDRC Confidential Questionnaire

DATE

MIDDLE

FIRST

NAME

LAST

1	31	61	91	121	151	181	211	241	271	301	331	361	391	421
2	32	62	92	122	152	182	212	242	272	302	332	362	392	422
3	33	63	93	123	153	183	213	243	273	303	333	363	393	423
4	34	64	94	124	154	184	214	244	274	304	334	364	394	424
5	35	65	95	125	155	185	215	245	275	305	335	365	395	425
6	36	66	96	126	156	186	216	246	276	306	336	366	396	426
7	37	67	97	127	157	187	217	247	277	307	337	367	397	427
8	38	68	98	128	158	188	218	248	278	308	338	368	398	428
9	39	69	99	129	159	189	219	249	279	309	339	369	399	429
10	40	70	100	130	160	190	220	250	280	310	340	370	400	430
11	41	71	101	131	161	191	221	251	281	311	341	371	401	431
12	42	72	102	132	162	192	222	252	282	312	342	372	402	432
13	43	73	103	133	163	193	223	253	283	313	343	373	403	433
14	44	74	104	134	164	194	224	254	284	314	344	374	404	434
15	45	75	105	135	165	195	225	255	285	315	345	375	405	435
BE SURE YOUR MARKS ARE HEAVY AND BLACK ERASE COMPLETELY ANY ANSWER YOU WISH TO CHANGE.														
16	46	76	106	136	166	196	226	256	286	316	346	376	406	436
17	47	77	107	137	167	197	227	257	287	317	347	377	407	437
18	48	78	108	138	168	198	228	258	288	318	348	378	408	438
19	49	79	109	139	169	199	229	259	289	319	349	379	409	439
20	50	80	110	140	170	200	230	260	290	320	350	380	410	440
21	51	81	111	141	171	201	231	261	291	321	351	381	411	441
22	52	82	112	142	172	202	232	262	292	322	352	382	412	442
23	53	83	113	143	173	203	233	263	293	323	353	383	413	443
24	54	84	114	144	174	204	234	264	294	324	354	384	414	444
25	55	85	115	145	175	205	235	265	295	325	355	385	415	445
26	56	86	116	146	176	206	236	266	296	326	356	386	416	446
27	57	87	117	147	177	207	237	267	297	327	357	387	417	447
28	58	88	118	148	178	208	238	268	298	328	358	388	418	448
29	59	89	119	149	179	209	239	269	299	329	359	389	419	449
30	60	90	120	150	180	210	240	270	300	330	360	390	420	450

E1

APPENDIX E

Standardization of the New London-NDRC

Confidential Questionnaire

SUMMARY

As a result of discussions in September of 1942 at the Submarine Base, it was decided that an inventory should be constructed which differed in form and content from those available at that time. Initiative for this project was taken by the New London unit of Project 44 on the suggestion of medical officers attached to the Submarine Base.

Preliminary discussions and revisions led to an inventory in two parts: part (1) of ten pages (286 items) and part (2) of six pages (172 items). In a typewritten form this inventory was first administered on October 22, 1942 to the crew of a submarine. In addition to the submarine crew, approximately 100 men drawn at random from the population of Submarine School candidates were also tested. From the distribution of the responses according to the psychiatric classifications of this sample population, it appeared that a large number of the questions might be expected to differentiate "Goods" and "Bads" reliably. Since this preliminary analysis showed promise, it was decided that the inventory research should be continued.

Testing experience had indicated some necessary changes, and a revision of the questions was undertaken.

Besides these changes, additional suggestions from different sources were incorporated in a second form. Altogether, this form comprised 112 questions, with 479 possible answers. Testing with the revision was started December 4, 1942 with 9 men from a second submarine. A large population of Submarine School candidates also was given the questionnaire, and on January 16, 1943 answer sheets for 356 Submarine School candidates were forwarded to the College Entrance Examination Board for item analysis. The object of this analysis was to determine whether there was any significant value in the second form, and if so, to derive a method for scoring the answer sheets.

PROCEDURE FOR EVALUATING THE FORM

The 356 Submarine School candidates were classified into two groups, 261 "Goods" and 95 "Bads". (The latter were recommended by a psychiatrist for disqualification; however, not all were finally disqualified by the Senior Medical Officer.) Men who subsequently failed Tank Training were listed as "Bads" regardless of preceding psychiatric judgment.

The data were transferred to standard IBM answer sheets, and a tabulation of responses for each of the two groups made by an IBM Test-Scoring machine. This tabulation shows that there are many items which tend to differentiate "Goods" and "Bads" with sufficient reliability to warrant their inclusion in an inventory. The Test-Scoring machine was also used to score the 356 answer sheets according to four scoring schemes.

METHOD OF ANALYSIS

Data for the selection of choices by two populations on a factual multiple-choice question are adapted to analysis by fourfold tables for chi square or to analysis by the Zubin nomographic method. The only caution that must be observed is that interpretations from such analysis are strictly limited to situations where precisely the same choices are presented, inasmuch as the selection of any particular choice is governed not only by the nature

of the choice, but also by the nature and applicability of the alternate choices. In other words, the distribution of responses is a function of the number and kind of the choices that are presented.

When instructions are given to make one choice and only one from a set of multiple factual choices, each choice may be considered one in a simple comparison. Thus, if choices 1, 2, 3, 4, and 5 are presented, one chooses either 1, or some one from the remaining complex of choices 2, 3, 4, and 5; likewise he chooses 2, or some one from the complex 1, 3, 4, and 5. Each choice, then, represents a simple forced alternate-choice situation. Either that item must be selected, or some one of the other possibilities must be selected instead.

Thus, for straightforward factual material, a multiple-choice question should yield approximately the same data as a combination of single-choice questions.

Let us take for an example the first question in the New London Inventory. The subject is asked to indicate one and only one of the following four items as the type of community in which he was brought up: (1) country, (2) village, (3) town, or (4) city. This single question should yield precisely the same data as the four single-choice questions as to whether one was brought up in: (1) (a) country or (b) village, town or city; (2) (a) village, or (b) country, town or city; (3) (a) town, or (b) country, village or city; and (4) (a) city or (b) country, village or town.

This argument will obtain so long as one is dealing with relatively factual material such that the suitability of the items is reasonably patent. With such factual material the response is governed by the phrasing of the question and by the number and nature of items included in the total number of choices. The response will be relatively independent of whether the choices are separate or clustered. On the other hand, the more "intangible" the material becomes for which choices are presented, the more the selection of the particular response will depend upon the factor of whether independent or clustered choices are presented.

For the most part, the New London Inventory is made up of questions about factual data of a man's life. We have therefore analyzed the results for the Inventory according to the following theoretical pattern. We are given two populations, G and B. For each question, individuals of these populations have exactly the same choices from which to select the most descriptive response. Suppose question  $\alpha$  to have 5 items, and that X persons from G population choose item (1), and Y from the B population choose the same item. Then (G-X) is the number of individuals from the G population who choose some one of items 2, 3, 4, and 5, and also (B-Y) is the number from the B population who choose some one from 2, 3, 4, and 5.

	G	B	Horizontal sums
1	x	y	x + y
2)			
3)			
Question $\alpha$	4) (G-x)	(B-y)	B+G - (x+y)
Vertical sums	5)		
	G	B	G + B

Thus we have a fourfold table for determining whether item (1) of question  $\alpha$  discriminates reliably between populations G and B. In a similar manner items 2, 3, 4, and 5 may be analyzed. The Fulcher-Zubin nomograph was used in this way to determine the reliability of all the items in the New London Inventory. Statistical stencils were constructed from the most reliable items as indicated by this method.

### RESULTS

Two of the scoring stencils were derived from analysis of the tabulation of results for the items. Another stencil was set up in an a priori manner, with no regard for the results in this tabulation. Still another consisted of a combination of the latter with one of the former. In the tables below, the results for this sample population can be seen for various cut-off scores for each of these four scoring devices. Of course, since the tabulation of the responses for this same population guided the construction of three of these stencils, these tables should not be used as a final basis for comparing the selective value of the four scoring schemes. Any such comparison should be based, instead, on the results for a population whose data are not involved in the selection of items to be included in the scoring stencils.

Scoring stencil #1 consists of items which the New London group believed might be expected, from an a priori viewpoint, to differentiate "Goods" and "Bads". This stencil rejects as follows:

TABLE E1  
Stencil #1

<u>Critical rejection score</u>	<u>Number re-jected from the 261 "Goods"</u>	<u>Percent-age</u>	<u>Number re-jected from the 95 "Bads"</u>	<u>Percent-age</u>
14	3	1	15	16
10	9	3	27	28
8	20	8	43	45
6	42	16	54	58
5	59	23	60	63
3	97	37	72	76

Stencil #3 consists of the items which differentiated with a high degree of reliability ( $P < 0.03$ ). The number of items scored (59) may be insufficient to yield an adequate range of scores. Rejection by this device is as follows:

TABLE E2  
Stencil #3

<u>Critical rejection score</u>	<u>Number rejected from the 261 "Goods"</u>	<u>Percent-age</u>	<u>Number rejected from the 95 "Bads"</u>	<u>Percent-age</u>
5	3	1	17	18
4	5	4	24	25
2	21	8	35	37
1	33	13	47	49
0	52	20	55	58
-2	82	31	71	75

Stencil #2 consists of a combination of the items on the #1 and #3 stencils. Rejection by this stencil is as follows:

TABLE E3  
Stencil #2

<u>Critical rejection score</u>	<u>Number rejected from the 261 "Goods"</u>	<u>Percent-age</u>	<u>Number rejected from the 95 "Bads"</u>	<u>Percent-age</u>
-1	3	1	15	16
-3	7	3	31	33
-5	16	6	47	49
-7	37	14	58	61
-8	46	18	68	72
-10	75	29	82	86

Finally, stencil #4 consists of all items which, for this sample differentiated between "Goods" and "Bads" with  $P < 0.10$ . In other words, #4 was a stencil of all items with a reliability exceeding a relatively low level.

TABLE E4  
Stencil #4

<u>Critical rejection score</u>	<u>Number rejected from the 261 "Goods"</u>	<u>Percent-age</u>	<u>Number rejected from the 95 "Bads"</u>	<u>Percent-age</u>
7	2	1	16	17
4	6	2	26	27
2	11	4	43	45
0	23	9	57	60
-1	33	12	63	66
-2	44	17	65	68
-4	65	25	70	74

DESCRIPTION OF STENCIL #4

Analyses carried out in this report have been based upon stencil #4.

The items marked (+) are those which indicate "badness" in terms of psychiatric criterion. Items marked (-) are indicative of "goodness". Thus a high (+) total score designates a man lying on the "bad" side of the distribution.

The table given below lists the differentiating items. In computing the total score, a point is added for each item marked (+) and a point subtracted for each item marked (-).

TABLE E5  
Differentiating Items

9+	32+	58+	75-	133-	169+	203+	233+	271-	313+
10-	36+	59+	76-	138-	172+	205-	234+	276+	314+
11+	37+	63-	78+	139+	173+	208-	237+	277+	324+
13+	39-	64+	87+	140+	174+	209+	238+	284+	325+
15+	43-	66+	89-	157+	176+	210+	241+	287-	333+
17+	45+	67+	98+	161+	177+	211-	242+	289+	336+
19+	46-	68+	102+	162+	178+	212+	248+	293+	337+
20-	51+	69-	104-	163+	179+	223-	251+	295+	346+
24+	52+	70+	106+	164+	187+	225+	250+	296+	352-
25-	53+	71-	113+	166+	188+	226+	262+	306+	357+
28+	54+	73+	115+	167+	189+	227+	264+	307+	
31-	56-	74+	131+	168+	200-	232+	268+	312+	

PERCENTILE CURVE

Percentile values have recently been determined (for stencil #4) on a population of 658 submarine men not used for standardizing purposes. These data are given in Table E6.

TABLE E6  
Reference Percentile Values (New Population)  
N=658.

<u>Score</u>	<u>Percentile</u>	<u>Score</u>	<u>Percentile</u>	<u>Score</u>	<u>Percentile</u>
-20		-8	44	4	96
-19	1	-7	50	5	97
-18	1	-6	56	6	98
-17	1	-5	62	7	99
-16	3	-4	68	8	99
-15	7	-3	74	9	99
-14	12	-2	77	10	99
-13	16	-1	82	11	99
-12	21	0	86	12	99
-11	27	1	89	13	99
-10	32	2	91	14	99
-9	38	3	93		

F1

APPENDIX F

Officers' Rating Scale

## MEDICAL RESEARCH DIVISION

U. S. SUBMARINE BASE

NEW LONDON CONNECTICUT

TO THE COMMANDING OFFICER:

THE PROBLEM

The problem of furnishing proper personnel for manning fleet submarines has become increasingly difficult due to rapid expansion, with resultant decreased periods for adequate training. Prior to the present emergency, preselection was effected in the fleet by requiring at least one year's sea service and recommendation by commanding officers, many of whom had the interests of the submarine service at heart. Now, the great majority of candidates for submarine school are either "boots" or men with less than six months service. Furthermore, many experienced fleet candidates now coming to us are not mentally or temperamentally suited for submarine service.

ATTEMPTED SOLUTION

At the direction of Commander Submarines, Atlantic, we are doing our best to cope with this problem. In collaboration with submarine commanding officers seeking improved personnel, a group of medical officers with years of submarine experience, working in conjunction with personnel specialists, are seriously attempting to evaluate candidates as to intelligence, emotional stability, reliability, neuromuscular coordination, and general aptitude for the strenuous and specialized duty of prolonged submarine patrol. These characteristics are evaluated by means of examinations, interviews, and an exhaustive battery of tests.

YOUR CONTRIBUTION

On the basis of the scores which

---

has made on this battery of tests, we have made certain predictions about this temperament and behavior. In order to evaluate this program it is absolutely essential to check these predictions. Therefore, the completion of the attached Performance Report is vital. It is realized that final judgment can be made by a superior officer only after observation of the man during a period of stress. So it is preferable that this report be returned to us after the completion of his first war patrol. Selection of proper personnel for submarines is considered of importance second to none. Without your assistance as a Commanding Officer, there is no method of checking on our efforts. With your help we can furnish you better men.

Ship \_\_\_\_\_ Date \_\_\_\_\_

Report covers period from \_\_\_\_\_ to \_\_\_\_\_

Report completed by \_\_\_\_\_

Comment \_\_\_\_\_

---

1. Adaptability to discipline: (willingness to obey orders; cooperativeness; attitude towards superiors.)

Insubordinate;  
disciplinary problem.

Surly, but not openly insubordinate.

Obeys without comment; does no more nor better than he must.

Obeys orders without comment. Unexceptionable but unremarkable performance of duty; appears willing most of time.

Obeys orders with dispatch and interest; willing at all times.

2. Intelligence: (exercise of good judgment; acuteness; - speed of understanding.)

Stupid; unable to get along; causes interference with work of others.

Bad judgment; dull but able to get along.

Judgment not always dependable; average intelligence.

Fair judgment; bright; catches on quickly.

Exercises good judgment; highly intelligent; understands readily.

3. Personal habits: (cleanliness of person, habits, dress, carriage.)

Immaculate in person, habits and dress; good carriage.

Usually clean; careless occasionally.

Presentable in appearance on most occasions.

Careless of appearance and gear.

Slovenly in appearance; dirty in habits; sloppy.

4. Use of alcohol.

Feetotaler.

Occasional drinker; rarely or never drunk.

Frequent drinker; occasionally drunk.

Heavy drinker; frequently drunk.

Chronic alcoholic or dipsomaniac; drunk whenever liquor available.

5. Dexterity.

Clumsy and inept with his hands.

Somewhat clumsy, but can learn to use hands fairly well.

Average ability to use hands.

Good with hands; fairly agile in manipulation.

Expert and agile; excellent with hands.

6. Alertness:  
(interest - initiative.)  
Always on toes; highly interested in work.  
Unusually alert; interested.  
Fairly interested; no initiative.  
Not very alert; uninterested.  
Dull; uninterested.

7. Courage:  
(reaction to stress).  
Unruffled by any danger; able to perform normally in all situations.  
Excited and upset by danger, but performs satisfactorily.  
Excited and upset by danger; work suffers slightly.  
Obvious fear, seriously interfering with performance of duty.  
Cowardly; goes to pieces under slightest stress.

8. Emotional maturity and stability.  
Very unstable; full of ups and downs; easily rattled.  
Occasionally rattled; somewhat moody; immature.  
Average  
Seldom shows signs of emotional disturbance.  
Stable, imperturbable, not bothered by new situations; never rattled.

9. Sociability:  
(ability to get along with other men.)  
Extremely well liked; good mixer; makes friends easily.  
Generally liked; few enemies.  
Liked quite well by most.  
Few friends; mixes poorly.  
Without friends; generally disliked and avoided.

10. Leadership:  
(degree to which respected, liked and followed).  
Ignored by others; always follows someone else's lead.  
Seldom acts as a leader; not well respected; ineffective in groups.  
Occasionally a leader; fairly well considered  
Usually leader of his own group.  
Lead followed by all; natural leader.

11.  
Tendency to  
create  
trouble.

Trouble-maker;  
frequently in  
fights; discipli-  
nary prob-  
lem.

Frequently in-  
volved in dis-  
turbance (thru  
his own fault).

Occasionally  
in some diffi-  
culty or other.

Seldom in diffi-  
culties over any-  
thing.

Never creates or  
is involved in  
difficulty with  
others.

12.  
Dependabil-  
ity: (res-  
ponsibility;  
reliability;  
industry).

Can be relied  
upon to per-  
form well under  
any circum-  
stances and at  
any time.

Reliable under  
most circum-  
stances; occa-  
sionally dis-  
appoints.

Average relia-  
bility.

Unreliable; can't  
be depended upon.

Completely un-  
dependable and  
irresponsible.

13.  
Adequacy of  
overall per-  
formance.

Will be trans-  
ferred off ship  
as soon as  
opportunity  
arises.

Will be trans-  
ferred if  
called upon for  
men.

Satisfactory.

Desire to retain.

Extremely valuable;  
special desire to  
retain.

Additional comment:

APPENDIX G

Correlations of the First 12 Items in  
the Officers' Rating Scale with the  
Final Item

Table G1 gives data on the correlations of each item of the first 12 items with the final item of the Officers' Rating Scale (Appendix F). It will be observed that all of the items correlate to a degree higher than .54 except item 4, "Use of Alcohol". The correlation for the "alcohol" item is only .36. Clearly, use of alcohol is not an important determiner of an officer's judgment.

The average score on the first 12 items has a correlation with item 13 of .92. Thus, the average score on the first 12 items and the "over-all" item score are highly related.

TABLE G1

Correlations (Uncorrected for Broad Categories) Between Each of the First 12 Items and the Final "Over-all" Item in Officers' Ratings

<u>Item</u>	<u>Correlation with Item 13</u>	<u>Item</u>	<u>Correlation with Item 13</u>
1	.69	7	.60
2	.80	8	.69
3	.69	9	.54
4	.36	10	.84
5	.66	11	.70
6	.84	12	.88
		Average Score on 12 Items	.92
		Average Score on Items 1, 3, 4, 7, 8, 9, 11, and 12	.90