

UNCLASSIFIED

AD NUMBER

ADA801521

CLASSIFICATION CHANGES

TO: UNCLASSIFIED

FROM: RESTRICTED

LIMITATION CHANGES

TO:
Approved for public release; distribution is unlimited.

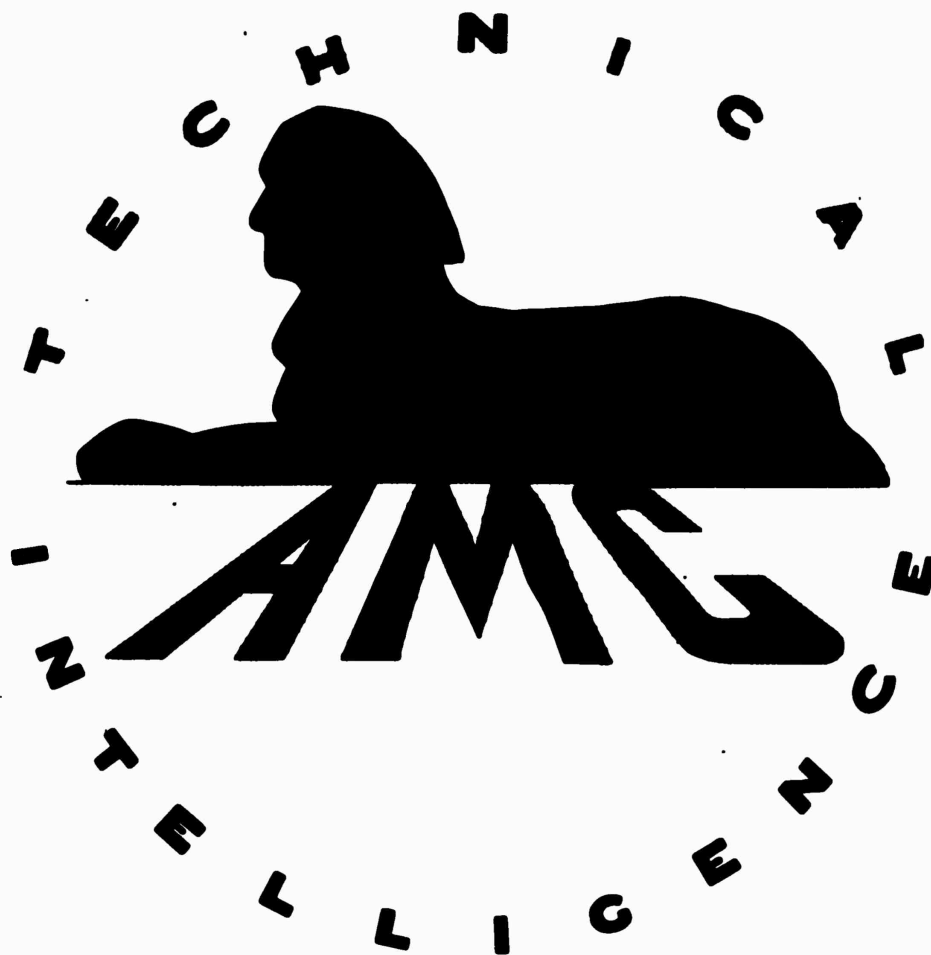
FROM:
Distribution authorized to U.S. Gov't. agencies and their contractors;
Administrative/Operational Use; 07 JUN 1943.
Other requests shall be referred to Office of Scientific Research and Development, Washington, DC 20301.

AUTHORITY

OSRD list no. 19 dtd 18-21 Feb 1946; OTS index dtd Jun 1947

THIS PAGE IS UNCLASSIFIED

Reproduced by
AIR DOCUMENTS DIVISION



HEADQUARTERS AIR MATERIEL COMMAND

WRIGHT FIELD, DAYTON, OHIO

The
U.S. GOVERNMENT

IS ABSOLVED

FROM ANY LITIGATION WHICH MAY
ENSUE FROM THE CONTRACTORS IN-
FRINGING ON THE FOREIGN PATENT
RIGHTS WHICH MAY BE INVOLVED.

WRIGHT FIELD

REEL - C

15000

A.T.I.

3 5 5 10

32 2
RESTRICTED

DECLASSIFIED
to
OPEN

**RESEARCH AND DEVELOPMENT OF THE
NAVY'S APTITUDE TESTING PROGRAM**

Under Contract with the
Office of Scientific Research and Development

**AVERAGES, STANDARD DEVIATIONS, AND
INTERCORRELATIONS OF NAVY
APTITUDE TESTS**

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

RESTRICTED

**RESEARCH AND DEVELOPMENT OF THE
NAVY'S APTITUDE TESTING PROGRAM**

**Harold Gulliksen, Director
Herbert S. Conrad
Norman Frederiksen**

**Report No. 4
June 7, 1943**

**AVERAGES, STANDARD DEVIATIONS, AND INTERCORRELATIONS
OF NAVY APTITUDE TESTS**

**Research Project under Contract
with the
Office of Scientific Research and Development
Contract OEMsr-705 Project N-106**

OSRD Report No. 1336

Copy No. 77 of 100 copies.

**Research and Statistical Laboratory
College Entrance Examination Board
Princeton, N. J.**

**This document contains information affecting the national
defense of the United States within the meaning of the
Espionage Act, U. S. C. 50; 31 and 32. Its transmission
or the revelation of its contents in any manner to an
unauthorized person is prohibited by law.**

This report is one of a series resulting from research contracts recommended to the National Defense Research Committee of the Office of Scientific Research and Development by the Committee on Service Personnel -- Selection and Training, National Research Council, Washington, D.C.

TABLE OF CONTENTS

	Page
Abstract	1
I. Introduction	3
II. Averages	4
III. Standard Deviations.	9
IV. Relation between School Grade Completed and Scores on the Tests	10
V. Intercorrelations.	28
VI. Frequency Distributions.	32
VII. Inconsistencies, Irregularities, Errors. . .	38
VIII. Summary.	48

**AVERAGES, STANDARD DEVIATIONS, AND INTERCORRELATIONS
OF NAVY APTITUDE TESTS**

ABSTRACT

The present report aims to answer some basic questions regarding (a) the quality of recruits at each of four large naval training stations, and (b) the aptitude tests used in the classification of recruits. The basis for the present report consists in data for 45,559 recruits, as given on Hollerith cards received from the Bureau of Naval Personnel. The principal findings are:

1. The men at the Great Lakes training station made significantly higher scores, on the average, than the men at Newport, San Diego, or Norfolk; on the average, the men at Norfolk made the lowest scores. From these facts it appears that the Great Lakes station could fill proportionately larger service-school quotas than the other stations. So long as significant differences among stations exist, it would appear desirable to make some adjustment in service-school quotas, in accord with observed differences; otherwise, a comparatively superior group of men is drawn from the high-ranking stations, and a comparatively inferior group from the low-ranking. The appropriate degree of quota-adjustment depends both on the extent of station-differences in the various tests (see Tables XII-XVI of the present report), and the relation between test scores and success in the service schools.

2. The Arithmetic and Mechanical Aptitude tests are of suitable difficulty; but the General Classification, English, and Spelling tests are too easy to provide the most efficient differentiation of superior from average or inferior recruits.

3. In general, men who make high scores on one test tend also to make better-than-average scores on the others. The relations among the various tests, however, are not uniform, some tests being more closely related than others. Specifically, the General Classification, English, and Spelling tests are fairly closely related to each other; the Mechanical Aptitude test shows, in general, a comparatively low relation with the other tests. The fairly close relations among the GCT, English, and Spelling tests suggest that these tests are all measuring more-or-less similar abilities; so that one or two of these tests could probably be eliminated from the general testing program without significant loss.

4. In general, the higher the school grade completed by a recruit, the higher his score on the aptitude tests; the increase in test scores is especially apparent between grades six and twelve, becoming smaller after grade twelve, and negligible after grade sixteen. These findings

support the common practice of favoring recruits who have at least a high-school education. By itself, however, the school grade completed by a recruit is not an adequate index of ability, since wide individual differences in test scores are found among men who are similar with respect to education.

5. The analysis of the data brought to light certain evidence of inconsistencies, irregularities, or errors. About five per cent of the Hollerith cards had to be rejected because of incompleteness (principally, lack of school-grade-completed), or because of obvious error (e.g., age of recruit reported as zero, or school-grade-completed reported as 95). Some of the test scores recorded on the Hollerith cards are impossible. The statistical data for the Mechanical Aptitude, Arithmetic, and English tests displayed noteworthy peculiarities at certain of the stations; thus, in Mechanical Aptitude, 28.2 per cent of the men at Norfolk were reported to have a score of 10; whereas less than one-half of one per cent of the men at Great Lakes were reported with this score. Such a huge discrepancy is probably due, in main part, not to a real difference between the men at the different stations, but to some inconsistency or irregularity of procedure. The correlations between the tests also differ excessively from station to station. All in all, it would appear that a single, given score cannot be depended upon to mean the same thing from person to person or station to station. The importance of these facts with regard to the selection of men for service schools is apparent.

The findings reported above confirm the conclusion reached in Report No. 1, namely, that variations in procedure from one station to another constitute a problem of fundamental importance and serious magnitude. It is for this reason that a highly systematic program of continuous coordination of procedures at the various stations is advisable. This suggests the desirability both of regularly scheduled visits to the stations by representatives from the Bureau of Naval Personnel, and of periodic statistical appraisal of test results from the various stations.

AVERAGES, STANDARD DEVIATIONS, AND INTERCORRELATIONS
OF NAVY APTITUDE TESTS

I. INTRODUCTION

The present study is based on the data in 45,559 Hollerith cards received from the Bureau of Naval Personnel, for the recruits reporting to Naval Training Stations between January and April, 1942. Each card contains the recruit's test scores and other information. The purposes of the present study are:

1. To compare the average quality of recruits received at each of four large naval training stations (Newport, Norfolk, Great Lakes, and San Diego).

2. To compare the individual differences (i.e., dispersion or variability) in quality of recruits received at each of the naval training stations.

3. To study the frequency distributions of scores in each test at each station, for such light as may be gained on validity and other matters.

4. To calculate the intercorrelations among the various tests at each station, and to interpret the results.

5. To determine the relation between school-grade-completed and scores on the various tests.

The specific test-scores included in the present study are:

1. Score on the O'Rourke General Classification Test, Junior Grade, Forms C and D, Navy Edition.

2. Score on the O'Rourke Mechanical Aptitude Test.

3. Score on U.S. Navy Standard Recruit Test in Arithmetic.

4. Score on U.S. Navy Standard Recruit Test in English.

5. Score on U.S. Navy Standard Recruit Test in Spelling.

II. AVERAGES

In the table on page 6 are presented the averages (means) and standard deviations of scores on each test at each station. The results for averages have been graphed in Chart I.

From Chart I it is clear that the Great Lakes station is receiving men who make higher scores, on the average, than any other station. The rank-order of each station on each test is given in Table I below.

TABLE I

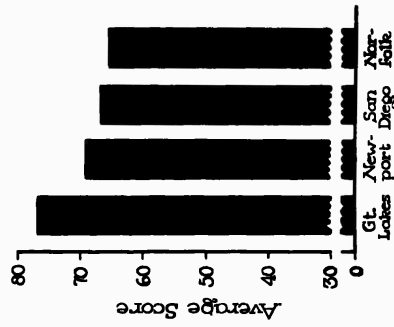
Rank of Each Station in Average Score on Each Test
(Based on 43,000 cards)

	Rank on				
	General Classification Test	Mechanical Aptitude Test	Arithmetic Test	English Test	Spelling Test
Great Lakes	1	1	1	1	1
Newport	2	5	2	2	5
San Diego	5	2	5	5	2
Norfolk	4	4	4	4	4

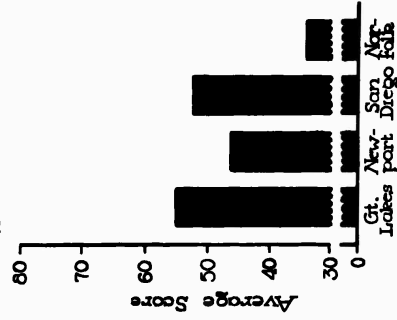
The essential point of Table I is that, on every test, the Great Lakes station ranks first, while the Norfolk station ranks last.

In connection with report No. 1, a total of about 1,000 answer sheets from Newport, Norfolk, Great Lakes, and San Diego were carefully rescored by the staff of the present project. It is safe to consider these data as relatively free from errors of scoring, recording, etc. For purposes of comparison, the data from these rescored papers have been presented in Table III. The averages in Table III give general support to those of Table II with respect to the relative positions of the Great Lakes and Norfolk stations; but in general, the amount of difference among the various stations is smaller in Table III. As an illustration, the difference between Great Lakes and Norfolk on the

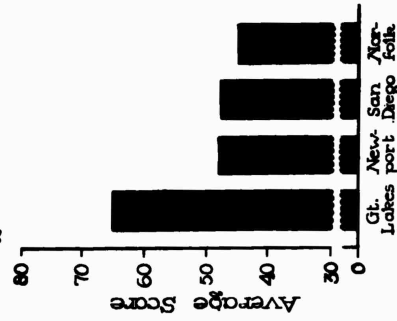
GENERAL CLASSIFICATION TEST
Average Score at Each Station



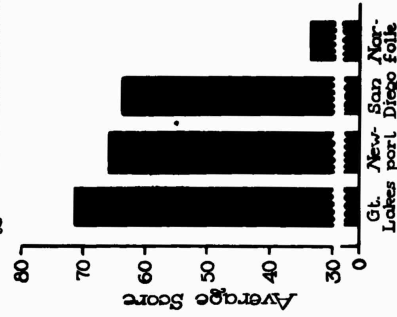
MECHANICAL APPTITUDE TEST
Average Score at Each Station



ARITHMETIC TEST
Average Score at Each Station



ENGLISH TEST
Average Score at Each Station



SPELLING TEST
Average Score at Each Station



CHART I. AVERAGE SCORE IN EACH TEST AT EACH NAVAL TRAINING STATION

TABLE II

Summary of Means and Standard Deviations of
Navy Aptitude Tests for All Stations
(Based on 43,000 cards)

Test by Stations	Number of Cases	Mean	Standard Deviation
GCT			
Newport	4858	69.35	14.53
Norfolk	11740	65.72	20.59
Great Lakes	15878	77.02	13.16
San Diego	8489	66.85	17.39
MAT			
Newport	4837	46.25	15.31
Norfolk	10798	33.83	25.86
Great Lakes	15650	55.04	27.27
San Diego	7866	51.54	16.82
Arithmetic			
Newport	4864	48.17	23.89
Norfolk	11753	44.75	26.80
Great Lakes	15820	65.03	23.45
San Diego	7808	47.83	25.00
English			
Newport	4852	65.81	12.63
Norfolk	11752	33.35	21.56
Great Lakes	15831	71.32	11.65
San Diego	7092	63.86	15.23
Spelling			
Newport	4832	57.13	20.57
Norfolk	11739	51.91	24.58
Great Lakes	15199	67.71	19.26
San Diego	7068	57.82	21.23

TABLE III

Means and Standard Deviations of
Navy Aptitude Tests for All Stations
(Based on Rescored Answer Sheets)

Tests by Stations	Number of Cases	Mean	Standard Deviation
GCT (Form B)			
Newport	204	71.18	15.14
Norfolk	298	65.51	17.84
GCT (Form C)			
Newport	239	69.42	15.04
Great Lakes	210	71.25	14.62
San Diego	210	66.35	17.61
LEAT*			
Newport	220	162.49	52.05
Norfolk	204	163.68	54.91
Great Lakes	210	165.46	53.93
San Diego	207	174.56	58.84
Arithmetic			
Newport	300	47.25	25.45
Norfolk	298	49.93	24.23
Great Lakes	301	63.67	23.29
San Diego	295	53.17	24.81
English			
Newport	300	68.38	12.62
Norfolk	298	63.45	14.95
Great Lakes	302	70.21	11.20
San Diego	297	69.74	12.75
Spelling			
Newport	300	58.27	22.21
Norfolk	298	56.51	21.36
Great Lakes	301	62.37	20.63
San Diego	296	58.50	21.30

* Data on this test are raw scores.

General Classification Test is 11.50 points in Table II, and only about half this, or 5.75 points, in Table III. In general, the Norfolk averages on the various tests in Table III run about $1/3$ to $1/2$ a standard deviation below those at Great Lakes.

III. STANDARD DEVIATIONS

In Table II on page 6 are presented data on the extent of individual differences in ability at each station, as measured by the standard deviation. Such data are of practical importance, because the greater the individual differences of ability within any station, the greater become the difficulties of successful instruction or training. From Table II it would appear that the Great Lakes station is, in general, characterized by the smallest spread of ability (lowest standard deviations), while Norfolk is characterized by the largest.* This conclusion should, however, be checked by reference to our data from rescored answer sheets, as compiled in Table III (page 7). The most impressive fact about the standard deviations in Table III, for each test respectively, is their comparative uniformity. To be sure, the standard deviations at Great Lakes are still somewhat smaller than at Norfolk, especially in GCT and in English; but in general, the differences among standard deviations in Table III are markedly less than in Table II. Since Table III is based upon highly accurate data, the conclusion is suggested that the much greater differences in Table II arise in large part from various inconsistencies of procedure at the various stations (see section VII). This interpretation seems more tenable than the other possible alternative, namely that the sample of approximately 1,000 rescored papers is highly unrepresentative. The extent of individual differences within the various stations may, for practical purposes, be considered reasonably uniform.

*The only exception to this occurs in the Mechanical Aptitude Test; and it seems certain that the results for this test are not fully comparable for all stations (see sections VI and VII below).

**IV. RELATION BETWEEN SCHOOL GRADE COMPLETED
AND SCORES ON THE TESTS**

One of the facts customarily employed in the selection of men is the educational grade attained; thus, the high school graduate is usually preferred to the man whose formal education stopped at the end of elementary school. To the extent that this educational standard is valid, a positive correlation should be found between school-grade-completed and aptitude test scores. An excessively high relationship between school-grade-completed and test scores, however, would raise question whether the tests reflect formal education too closely, or duplicate excessively the information already provided by school-grade-completed.

The facts on the relationship between school grade and test scores, based on the 45,000 cards supplied by the Bureau of Naval Personnel, are given in detail in Tables V-IX. In these tables may be found, for each test, the average and the standard deviation of scores for men completing each grade of school, respectively; the data are presented for each station separately, and also for all stations combined. For all stations combined, the correlation-ratios between school-grade-completed and test scores are as set forth in Table IV below.

TABLE IV

Correlation-Ratio between School Grade Completed
and Test Scores

Test	Correlation Ratio
General Classification	.55
Mechanical Aptitude	.58
Arithmetic	.51
English	.48
Spelling	.52

None of the correlation-ratios in Table IV is high enough to justify substitution of school-grade-completed for any of the tests -- particularly in view of the case of falsification of the information for school-grade-completed. The correlation-ratio is highest for the General Classification test, and lowest for Mechanical Aptitude. This accords with the fact that general verbal intelligence is definitely more closely related to success and continuation in school, than is mechanical aptitude.

Chart II presents, for the total sample (all stations combined), a graphical summary of the relation between school-grade-completed and test scores. For every test, between grades six and twelve, there is on the average a definite advance of score with increase of education; from grade twelve to grade sixteen the increase tends to be smaller (nil, in the case of the Mechanical Aptitude test); beyond grade sixteen (college graduation), the average change is negligible. From Chart II it appears that the men reported to have education of less than fifth grade make slightly superior test scores than those who have had only fifth-grade education.*

It should perhaps be emphasized that, among men within any given educational classification, wide individual differences in test scores occur. This is indicated in the graphs of Chart II by the lines labelled + 1 S.D. and - 1 S.D. ("S.D." being the standard deviation of scores at each educational classification). The same fact was brought out by the correlation-ratios previously presented: a correlation-ratio of .50, for example, means that the dispersion of scores made by men of a given educational classification is, on the average, about 85% as large as the dispersion of scores made by the total group (including all educational levels). It is important not to misinterpret Chart II as indicating a strong relation between school-grade-completed and test scores.

*This may very well represent the true situation. On the other hand, it is possible that a certain proportion of those recorded with "0" in education merely failed to state their last grade completed; such persons, on the average, may have had higher than fifth-grade education.

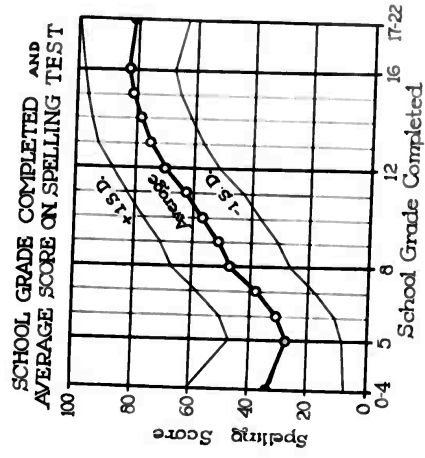
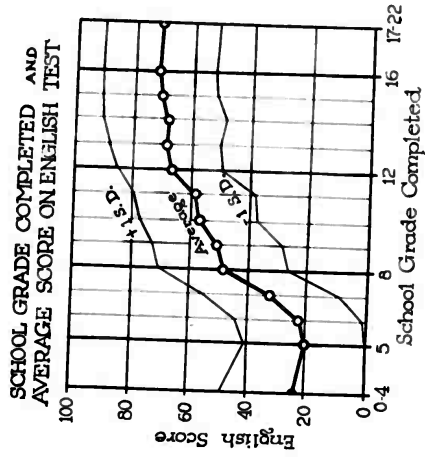
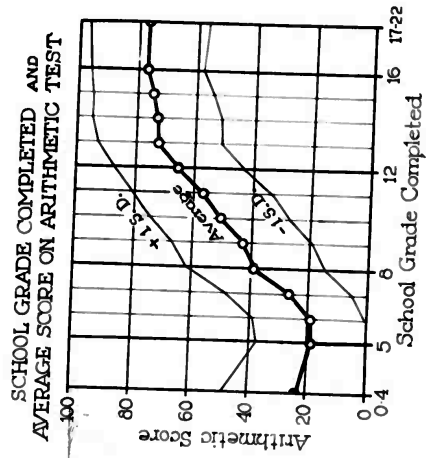
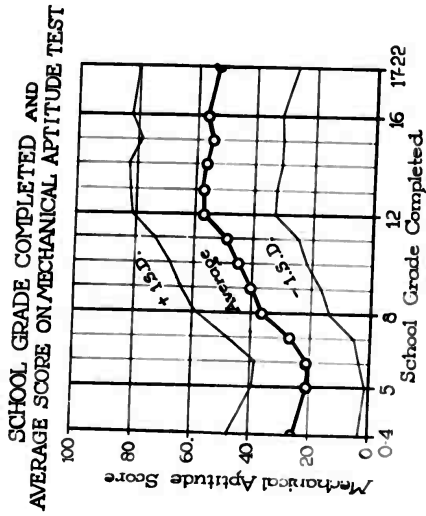
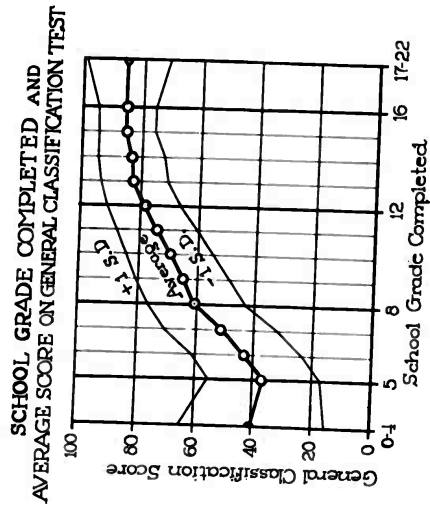


CHART II. RELATION BETWEEN SCHOOL GRADE COMPLETED AND AVERAGE SCORE ON EACH TEST
(ALL TRAINING STATIONS COMBINED)

TABLE V
Means and Standard Deviations for the
General Classification Test by Grade Completed

School Grade Completed	Newport		
	Number of Cases	Mean	Standard Deviation
0 - 4	12	55.58	19.99
5	7	44.71	22.41
6	58	49.07	15.44
7	184	55.23	14.45
8	801	60.19	14.15
9	719	64.78	12.73
10	994	69.48	12.31
11	641	73.26	12.19
12	1289	77.39	10.98
13	60	77.95	13.63
14	43	80.60	10.22
15	16	81.81	7.85
16	23	83.26	11.17
17 - 22	11	82.64	13.97
Total	4858	69.35	14.53

School Grade Completed	Norfolk		
	Number of Cases	Mean	Standard Deviation
0 - 4	105	37.13	23.99
5	152	34.02	19.08
6	461	40.13	19.01
7	999	49.14	19.85
8	1558	55.44	18.04
9	1427	60.71	17.67
10	1510	65.58	16.73
11	1947	73.03	15.59
12	2488	76.63	14.94
13	441	80.74	13.07
14	329	81.95	13.24
15	127	85.23	11.81
16	146	85.08	8.95
17 - 22	50	84.88	16.49
Total	11740	65.72	20.59

TABLE V - Continued

Means and Standard Deviations for the
General Classification Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Great Lakes			
0 - 4	9	63.22	19.81
5	14	56.57	18.96
6	76	55.43	17.66
7	309	58.87	17.86
8	2212	65.64	14.69
9	1531	71.09	12.80
10	2355	74.72	11.62
11	1998	78.54	10.10
12	6059	81.88	9.74
13	636	85.97	7.99
14	359	87.06	6.91
15	141	87.62	6.37
16	144	87.19	9.59
17 - 22	35	89.63	7.33
Total	15878	77.02	13.16
San Diego			
0 - 4	21	39.86	27.57
5	33	42.82	19.85
6	113	39.07	17.26
7	297	46.99	18.01
8	1076	53.90	16.66
9	849	60.43	14.99
10	1213	64.26	15.69
11	1100	68.33	14.64
12	3018	73.91	13.27
13	397	78.83	12.00
14	187	79.76	12.42
15	83	82.42	9.12
16	85	82.62	9.54
17 - 22	17	79.29	15.13
Total	8489	66.85	17.39

TABLE V - Concluded

Means and Standard Deviations for the
General Classification Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
All Stations			
0 - 4	147	40.63	25.19
5	206	37.33	20.33
6	708	42.34	19.11
7	1789	51.09	19.17
8	5647	60.21	16.90
9	4526	64.82	15.62
10	6072	69.50	14.66
11	5686	74.08	13.83
12	12854	78.54	12.35
13	1534	82.31	11.42
14	918	83.44	11.26
15	367	85.37	9.48
16	398	85.39	9.60
17 - 22	113	85.29	14.23
Total	40965	70.77	17.38

TABLE VI

Means and Standard Deviations for the
Mechanical Aptitude Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Newport			
0 - 4	12	36.83	12.99
5	7	35.14	16.57
6	57	36.09	14.79
7	183	37.09	13.80
8	798	41.08	14.70
9	721	43.08	13.74
10	990	46.16	14.41
11	636	49.04	15.01
12	1283	51.15	15.13
13	60	50.62	17.38
14	41	52.00	14.34
15	16	55.06	16.05
16	23	54.74	13.32
17 - 22	10	51.10	16.84
Total	4837	46.25	15.31
Norfolk			
0 - 4	63	23.52	23.97
5	101	14.37	13.23
6	338	15.11	13.52
7	839	20.02	18.37
8	1110	23.30	20.42
9	1296	28.60	23.12
10	1400	32.24	23.72
11	1861	38.24	25.45
12	2416	44.03	27.54
13	434	41.26	27.36
14	323	41.18	26.51
15	125	42.16	25.53
16	143	47.42	27.53
17 - 22	49	46.65	29.71
Total	10798	33.83	25.86

TABLE VI - Continued

Means and Standard Deviations for the
Mechanical Aptitude Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Great Lakes			
0 - 4	9	20.00	28.89
5	14	26.43	29.28
6	73	21.56	26.04
7	302	20.95	28.13
8	2186	39.92	27.82
9	1507	45.21	27.30
10	2320	50.46	25.92
11	1964	56.54	24.94
12	5974	63.72	23.80
13	631	66.56	23.50
14	356	67.22	23.86
15	140	64.48	24.08
16	139	65.01	24.94
17 - 22	35	59.51	26.97
Total	15650	55.04	27.27
San Diego			
0 - 4	19	29.89	17.20
5	28	36.61	19.43
6	102	30.12	12.66
7	267	35.43	14.60
8	969	42.67	15.79
9	778	46.67	16.18
10	1107	49.32	15.25
11	998	52.75	15.22
12	2857	56.97	15.31
13	384	59.58	15.31
14	180	58.33	15.68
15	80	57.95	14.94
16	80	58.32	15.47
17 - 22	17	60.71	15.16
Total	7866	51.54	16.82

TABLE VI - Concluded

Means and Standard Deviations for the
Mechanical Aptitude Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
All Stations			
0 - 4	103	25.94	22.85
5	150	20.61	19.19
6	570	20.72	17.43
7	1591	26.46	20.89
8	5363	36.22	23.73
9	4302	40.11	23.65
10	5817	45.13	23.19
11	5459	48.73	23.99
12	12530	57.10	23.47
13	1509	57.69	24.60
14	900	56.48	25.25
15	361	54.89	24.60
16	385	56.47	24.98
17 - 22	111	53.26	26.84
Total	39151	47.40	25.40

TABLE X

Means and Standard Deviations for the
Arithmetic Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Newport			
0 - 4	12	36.67	22.85
5	7	29.43	23.49
6	58	19.22	18.61
7	184	27.83	19.07
8	802	34.91	20.55
9	721	40.76	21.44
10	997	47.82	22.08
11	640	54.75	21.91
12	1291	59.53	21.17
13	60	67.37	20.59
14	43	67.28	21.18
15	16	69.00	22.36
16	23	73.91	18.88
17 - 22	10	73.90	24.47
	4864	48.17	23.89
Norfolk			
0 - 4	106	20.55	24.76
5	152	15.97	17.25
6	464	18.14	16.69
7	997	23.71	19.13
8	1566	30.48	20.80
9	1427	37.06	22.58
10	1511	43.08	22.81
11	1947	52.67	23.85
12	2491	58.68	23.47
13	442	66.05	22.86
14	329	66.69	24.22
15	127	72.19	21.53
16	146	74.72	20.22
17 - 22	48	76.54	19.61
Total	11753	44.75	26.80

TABLE VII - Continued
 Means and Standard Deviations for the
 Arithmetic Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Great Lakes			
0 - 4	9	35.56	22.42
5	14	35.71	25.83
6	75	34.40	28.57
7	310	39.99	26.95
8	2200	48.92	24.52
9	1520	53.84	23.42
10	2341	61.38	21.84
11	1998	65.66	20.69
12	6039	73.15	18.68
13	638	80.04	16.85
14	357	83.18	15.09
15	141	81.74	15.58
16	143	82.23	15.63
17 - 22	35	80.80	15.37
Total	15820	65.03	23.45
San Diego			
0 - 4	19	28.68	21.08
5	27	20.56	18.87
6	101	15.85	14.18
7	261	24.73	19.27
8	958	29.05	19.38
9	773	36.54	20.21
10	1095	41.71	21.53
11	989	49.07	22.75
12	2849	57.35	22.04
13	381	67.72	20.80
14	179	69.12	22.05
15	79	68.22	22.64
16	80	73.21	19.54
17 - 22	17	71.41	24.77
Total	7808	47.83	25.00

TABLE VII - Continued
 Means and Standard Deviations for the
 Arithmetic Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Great Lakes			
0 - 4	9	35.56	22.42
5	14	35.71	25.83
6	75	34.40	28.57
7	310	39.99	26.95
8	2200	48.92	24.52
9	1520	53.84	23.42
10	2341	61.38	21.84
11	1998	65.66	20.69
12	6039	73.15	18.68
13	638	80.04	16.85
14	357	83.18	15.09
15	141	81.74	15.58
16	143	82.23	15.63
17 - 22	35	80.80	15.37
Total	15820	65.03	23.45
San Diego			
0 - 4	19	28.68	21.08
5	27	20.56	18.87
6	101	15.85	14.18
7	261	24.73	19.27
8	958	29.05	19.38
9	773	36.54	20.21
10	1095	41.71	21.53
11	989	49.07	22.75
12	2849	57.35	22.04
13	381	67.72	20.80
14	179	69.12	22.05
15	79	68.22	22.64
16	80	73.21	19.54
17 - 22	17	71.41	24.77
Total	7808	47.83	25.00

TABLE VII - Concluded

Means and Standard Deviations for the
Arithmetic Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
All Stations			
0 - 4	146	23.86	24.69
5	200	18.44	19.25
6	698	19.65	18.91
7	1752	27.18	21.62
8	5526	38.22	23.81
9	4441	43.31	23.60
10	5944	50.83	23.73
11	5574	56.93	23.33
12	12670	65.36	22.02
13	1521	72.39	20.95
14	908	73.68	21.87
15	363	74.90	20.54
16	392	77.10	18.86
17 - 22	110	76.86	20.07
Total	40245	53.73	26.47

TABLE VIII

Means and Standard Deviations for the
English Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Newport			
0 - 4	12	57.67	23.12
5	7	31.43	26.13
6	58	49.26	15.49
7	184	53.10	14.43
8	797	59.16	11.96
9	721	62.47	10.66
10	993	65.63	10.90
11	640	69.69	10.29
12	1288	71.70	10.56
13	60	71.95	11.33
14	43	75.56	9.87
15	16	76.62	5.98
16	23	82.17	11.45
17 - 22	10	77.90	11.18
Total	4852	65.81	12.63
Norfolk			
0 - 4	106	15.32	19.42
5	152	13.16	16.16
6	464	12.17	12.21
7	1000	17.65	14.71
8	1563	21.92	16.03
9	1427	26.65	16.82
10	1512	31.52	17.82
11	1949	39.55	19.31
12	2489	44.19	20.02
13	441	50.39	20.18
14	329	52.42	20.86
15	126	56.17	21.22
16	146	57.77	18.70
17 - 22	48	63.25	22.70
Total	11752	33.35	21.56

TABLE VIII - Continued

Means and Standard Deviations for the
English Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Great Lakes			
0 - 4	9	55.56	13.75
5	14	53.14	12.19
6	75	52.24	16.17
7	311	57.72	14.51
8	2203	62.54	11.45
9	1519	65.94	11.05
10	2344	68.88	10.30
11	1998	72.00	9.91
12	6045	75.62	9.04
13	636	79.50	9.08
14	358	80.91	8.05
15	141	80.48	9.92
16	143	82.88	9.21
17 - 22	35	82.06	8.74
Total	15831	71.32	11.65
San Diego			
0 - 4	16	39.06	20.28
5	23	42.65	13.43
6	83	38.00	18.55
7	222	48.29	16.77
8	826	52.01	14.50
9	679	58.31	13.37
10	983	61.53	13.36
11	901	64.62	13.06
12	2658	69.15	12.21
13	363	74.15	10.81
14	171	75.08	12.04
15	73	75.63	10.36
16	77	76.14	12.56
17 - 22	17	69.00	19.01
Total	7092	63.86	15.23

TABLE VIII - Concluded

Means and Standard Deviations for the
English Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
All Stations			
0 - 4	143	24.06	24.92
5	196	20.13	20.94
6	680	22.91	21.30
7	1717	32.67	23.33
8	5389	48.64	22.04
9	4346	51.27	22.03
10	5832	57.40	20.40
11	5488	58.99	20.57
12	12480	67.57	17.48
13	1500	69.35	18.53
14	901	69.14	19.56
15	356	70.71	18.47
16	389	72.08	18.18
17 - 22	110	71.45	19.65
Total	39527	58.02	22.78

TABLE IX

Means and Standard Deviations for the
Spelling Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Newport			
0 - 4	12	60.17	25.13
5	7	53.14	18.60
6	56	38.21	21.80
7	184	39.65	17.56
8	795	45.26	18.93
9	719	50.53	18.31
10	989	56.28	18.25
11	634	61.08	18.06
12	1284	68.25	17.43
13	60	68.73	20.59
14	43	74.33	16.63
15	16	79.62	18.06
16	23	83.83	18.73
17 - 22	10	79.60	21.12
Total	4832	57.13	20.57
Norfolk			
0 - 4	105	29.20	24.67
5	152	24.15	17.69
6	463	27.22	18.02
7	999	35.23	19.47
8	1562	39.50	19.99
9	1427	44.85	20.90
10	1509	49.83	21.44
11	1947	56.77	21.72
12	2486	64.08	21.23
13	440	69.35	21.03
14	328	73.39	19.92
15	127	79.15	18.52
16	146	79.68	16.30
17 - 22	48	82.46	17.51
Total	11739	51.91	24.58

TABLE IX - Continued
 Means and Standard Deviations for the
 Spelling Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
Great Lakes			
0 - 4	9	52.00	27.34
5	13	40.77	19.78
6	75	43.59	23.46
7	310	47.10	21.26
8	2176	54.69	19.70
9	1496	58.57	18.90
10	2280	63.49	17.92
11	1942	69.01	16.54
12	5686	74.56	15.33
13	595	81.28	13.19
14	323	84.63	10.99
15	131	84.59	10.56
16	134	86.71	11.63
17 - 22	29	84.69	12.35
Total	15199	67.71	19.26
San Diego			
0 - 4	16	34.62	20.54
5	23	38.87	19.20
6	82	30.52	16.41
7	222	36.33	19.26
8	821	42.37	18.05
9	675	46.90	18.13
10	979	52.54	18.44
11	901	57.95	18.07
12	2649	65.50	18.31
13	363	73.28	16.97
14	170	77.62	15.36
15	73	78.37	15.00
16	77	81.49	15.50
17 - 22	17	74.35	24.21
Total	7068	57.82	21.23

TABLE IX - Concluded

Means and Standard Deviations for the
Spelling Test by Grade Completed

School Grade Completed	Number of Cases	Mean	Standard Deviation
All Stations			
0 - 4	142	33.87	26.54
5	195	27.32	18.97
6	676	30.35	19.67
7	1715	37.99	20.09
8	5354	46.97	20.53
9	4317	50.87	20.26
10	5757	56.81	19.91
11	5424	62.57	19.58
12	12105	69.75	18.19
13	1458	75.17	17.94
14	864	78.47	16.76
15	347	81.06	15.42
16	380	82.78	15.14
17 - 22	104	81.48	16.30
Total	38838	59.82	22.53

V. INTERCORRELATIONS

The correlations among the various tests are of interest in several respects. We wish to know, for example, whether any of the tests is so highly correlated with any other as to constitute a virtual duplication. We wish to know whether the General Classification test shows (in conformity with its title) the highest average correlation with other tests. We wish to know if the correlations among tests differ markedly among the different stations.

In order to determine the correlations among the various tests, a sample of 2,000 cards -- 500 from each naval training station -- was drawn. The procedure employed to assure an unbiased sample from the total 45,000 cards was as follows:

1. The 45,000 cards were classified into four groups, one for each naval training station.
2. The cards for each station were then classified into fourteen groups according to school grade completed.
3. A fixed percentage of the cards in each group was selected at random* to comprise a sample of about 800 cards for each station.
4. The cards thus selected were next checked for obvious errors -- specifically, for (a) duplicate names; (b) duplicate service numbers; (c) complete duplication of scores on all tests; and (d) impossible test scores (see section VII, b).
5. From the cards remaining, enough cards were discarded at random* to reduce the total to exactly 500 for each station.

The intercorrelations for the full sample of 2,000 cases (all stations combined) are given in Table X; and for each station separately, in Table XI. From Table X it is clear that the General Classification test shows the highest average correlation with other tests; the Mechanical Aptitude test shows the lowest. Several of the intercorrelations are over .50, and that between General Classification test and Spelling is over .60. Taken as they stand, the intercorrelations are high enough to raise question whether one or two tests could not be omitted without significant loss in the multiple correlation with a suitable criterion. It may deserve notice that the r 's in both Tables X and XI would in

*The "random" method consisted simply in selecting or withdrawing every n 'th card -- n being adjusted, in each case, to yield the sample of desired size.

general be somewhat higher if all differences in circumstances of administration, or errors of scoring, coding, etc. were eliminated.

The intercorrelations among the tests at each station are given in Table XI. Considerable variation among stations may be observed. Thus the correlation between General Classification test and Mechanical Aptitude test at Great Lakes is .42, at San Diego is .65; the correlation between General Classification test and Arithmetic at Great Lakes is .49, at Norfolk is .61; etc. In general, the correlations are comparatively high at San Diego: the average of all the correlations at San Diego is .56, at the other stations, .45.* Some of the largest discrepancies in correlations are observed in connection with the Mechanical Aptitude test; some of the smallest, in connection with the Spelling test. By no means all the observed variation of correlations among the stations can be ascribed to chance.**

*The simple averaging of correlation coefficients is the most convenient, even if not theoretically the most valid, procedure. Use of a more elaborate and theoretically more valid procedure would not alter our results significantly.

**For a discussion of the problem of heterogeneity of relationships among tests at the different stations, see section VII, e.

TABLE X

Intercorrelations of Navy Aptitude Tests
(All Stations Combined, N = 2,000)

	General Classification Test	Mechanical Aptitude Test	Arithmetic Test	English Test	Spelling Test
GCT	---	.48	.58	.52	.61
MAT	.48	---	.58	.48	.54
Arithmetic	.58	.58	---	.44	.55
English	.52	.48	.44	---	.55
Spelling	.61	.34	.55	.55	---
Average	.55	.42	.48	.49	.50

TABLE XI

Intercorrelations of Navy Aptitude Tests
(N = 500 for Each Station)

	GCT	MAT	Arith- metic	English	Spell- ing	Average
GCT						
Newport....	---	.48	.51	.58	.58	.55
Norfolk....	---	.50	.61	.64	.60	.59
Great Lakes	---	.42	.49	.54	.55	.50
San Diego..	---	.65	.56	.69	.66	.64
MAT						
Newport....	.46	---	.29	.29	.22	.32
Norfolk....	.50	---	.59	.52	.30	.58
Great Lakes	.42	---	.27	.29	.22	.30
San Diego..	.65	---	.47	.55	.39	.51
ARITHMETIC						
Newport....	.51	.29	---	.46	.44	.45
Norfolk....	.61	.59	---	.49	.49	.50
Great Lakes	.49	.27	---	.48	.52	.44
San Diego..	.56	.47	---	.57	.55	.54
ENGLISH						
Newport....	.58	.29	.46	---	.54	.47
Norfolk....	.64	.52	.49	---	.55	.50
Great Lakes	.54	.29	.48	---	.56	.47
San Diego..	.69	.55	.57	---	.59	.60
SPELLING						
Newport....	.58	.22	.44	.54	---	.45
Norfolk....	.60	.30	.49	.55	---	.49
Great Lakes	.55	.22	.52	.56	---	.48
San Diego..	.66	.39	.55	.59	---	.55

omit

VI. FREQUENCY DISTRIBUTIONS OF SCORES ON EACH TEST

The chief function of any test is to provide a range of scores within which individual differences in the ability measured by the test may be distinguished. It is especially important that there should be no undue "piling up" of cases at any particular score. It is reasonable, for example, to expect a heavy concentration of scores at or near the middle of the distribution; in no test, however, should there be a considerable concentration at either very high or very low scores (the term "low" referring to scores which are near or below the level that could be achieved by random or chance responses to the test). For the single, specific purpose of differentiating those who should be sent to service school from those who should not, it is desirable that the test be of such difficulty that the service school men make scores at or above the "50 per cent point" (i.e., 50 per cent of a perfect score). The proportion of men sent to service schools, however, varies from time to time with the available manpower and the needs of the Navy; so that for general purposes it is best that the fifty-per-cent score be made by the typical or "average" man.

The frequency distributions of scores in the five tests are given, for each station separately, and for the four stations combined, in Tables XII-XVI. In each distribution, the range of scores is clearly adequate to provide good differentiation of individual differences. In other respects, however, the distributions are subject to question. In particular, the General Classification, English, and Spelling tests appear too easy to provide the most efficient differentiation -- both for general purposes, and especially for the specific purpose of selecting men for service schools. The Arithmetic and Mechanical Aptitude tests are of suitable average difficulty, but do not exhibit as much concentration of cases at the middle as one would normally expect. If the scores from the individual stations are examined closely, certain anomalies appear -- e.g., the heavy concentration of high scores in Arithmetic at the Great Lakes station, the large number of zero scores recorded for the English test at Norfolk, etc. These features, however, appear to arise from inconsistencies or irregularities of procedure at some of the stations, rather than from features characteristic of the particular tests; these matters are considered in some detail in the following section of the present report.

55
TABLE XII

Distributions of Scores on General Classification Test

Scores	Newport Freq.	Norfolk Freq.	Great Lakes Freq.	San Diego Freq.	Total Freq.
95-99	50	388	534	139	1111
90-94	235	779	1856	454	3324
85-89	462	1118	2807	726	5113
80-84	565	1222	2731	954	5472
75-79	666	1208	2362	1011	5247
70-74	683	1094	1856	941	4573
65-69	534	1025	1292	845	3696
60-64	492	896	951	802	3141
55-59	364	859	506	684	2413
50-54	345	748	362	574	2029
45-49	181	629	221	407	1338
40-44	134	488	158	335	1115
35-39	73	313	108	216	710
30-34	37	296	37	130	500
25-29	17	236	37	96	386
20-24	7	202	22	79	310
15-19	5	112	10	42	169
10-14	4	113	10	30	157
5-9	1	59	8	11	79
0-4	3	55	11	13	82
Total	4858	11740	15878	8489	40965
Mean*	69.35	65.72	77.02	66.85	70.77
Standard Deviation*	14.53	20.59	13.16	17.39	17.38

* The means and standard deviations are computed from ungrouped data, not from the frequency distributions.

TABLE XIII

Distributions of Scores on Mechanical Aptitude Test

Scores	Newport Freq.	Norfolk Freq.	Great Lakes Freq.	San Diego Freq.	Total Freq.
95-99		157	345		
90-94	9	178	880	5	507
85-89	29	255	1118	14	1081
80-84	60	272	1164	93	1495
75-79	99	255	1040	236	1732
70-74	148	283	1155	378	1772
65-69	267	344	1120	529	2115
60-64	321	438	1137	627	2358
55-59	473	448	1041	780	2676
50-54	514	392	768	796	2758
45-49	626	284	508	883	2557
40-44	648	520	756	823	2241
35-39	547	471	631	741	2665
30-34	453	576	676	585	2234
25-29	296	531	654	449	2156
20-24	185	693	581	398	1879
15-19	102	767	567	355	1814
10-14	37	3660	376	146	1582
5- 9	10	224	8	5	4078
0- 4	13	48	1125	8	250
Total	4837	10798	15650	7866	39151
Mean*	46.25	33.83	55.04	51.54	47.40
Standard Deviation*	15.31	25.86	27.27	16.82	25.40

* The means and standard deviations are computed from ungrouped data, not from the frequency distributions.

TABLE XIV

Distributions of Scores on Arithmetic Test

Scores	Newport Freq.	Norfolk Freq.	Great Lakes Freq.	San Diego Freq.	Total Freq.
95-99	115	390	1625	262	2392
90-94	126	356	1281	198	1961
85-89	185	442	1409	311	2347
80-84	179	482	1397	315	2373
75-79	271	540	1374	402	2587
70-74	307	560	1272	400	2539
65-69	286	636	1084	521	2527
60-64	325	606	1039	466	2436
55-59	325	648	942	542	2457
50-54	338	736	798	509	2381
45-49	359	584	700	541	2184
40-44	340	692	610	444	2086
35-39	305	662	503	499	1969
30-34	275	716	455	469	1915
25-29	290	645	356	478	1769
20-24	266	778	309	424	1777
15-19	219	709	251	362	1541
10-14	161	642	177	288	1268
5- 9	185	482	130	199	996
0- 4	7	447	108	178	740
Total	4864	11753	15820	7808	40246
Mean*	48.17	44.75	65.03	47.83	53.73
Standard Deviation*	23.89	26.80	23.45	25.00	26.47

* The means and standard deviations are computed from ungrouped data, not from the frequency distributions.

TABLE XV

Distributions of Scores on English Test

Scores	Newport Freq.	Norfolk Freq.	Great Lakes Freq.	San Diego Freq.	Total Freq.
95-99	7	10	72	6	95
90-94	49	43	466	101	659
85-89	162	85	1093	276	1616
80-84	365	159	2096	515	3135
75-79	538	194	2991	852	4475
70-74	884	294	3152	1118	5448
65-69	903	330	2497	1062	4792
60-64	752	486	1588	967	3793
55-59	466	529	804	605	2404
50-54	275	641	461	462	1839
45-49	152	687	265	312	1416
40-44	107	910	167	265	1449
35-39	88	870	105	197	1260
30-34	43	979	85	141	1248
25-29	20	939	31	93	1083
20-24	16	1066	22	42	1136
15-19	9	949	14	29	1001
10-14	8	887	7	12	914
5-9	4	721	4	16	745
0-4	4	977	11	21	1019
Total	4852	11752	15831	7092	39527
Mean*	65.81	33.35	71.32	63.86	58.02
Standard Deviation*	12.63	21.56	11.65	15.23	22.78

* The means and standard deviations are computed from ungrouped data, not from the frequency distributions.

TABLE XVI

Distributions of Scores on Spelling Test

Scores	Newport Freq.	Norfolk Freq.	Great Lakes Freq.	San Diego Freq.	Total Freq.
95-99	99	341	769	191	1400
90-94	198	464	1276	315	2252
85-89	162	419	1152	301	2034
80-84	380	692	1771	479	3322
75-79	227	548	1279	425	2479
70-74	460	837	1708	641	3646
65-69	274	566	1116	448	2404
60-64	516	963	1391	670	3540
55-59	283	596	843	438	2160
50-54	473	898	1102	676	3149
45-49	298	632	610	417	1957
40-44	479	909	808	573	2769
35-39	218	594	416	353	1581
30-34	314	791	467	438	2010
25-29	156	503	196	249	1104
20-24	161	728	157	243	1289
15-19	63	429	66	83	641
10-14	46	460	27	73	606
5-9	11	228	19	27	285
0-4	14	141	27	28	210
Total	4832	11739	15199	7068	38838
Mean*	57.13	51.91	67.71	57.82	59.82
Standard Deviation*	20.57	24.58	19.26	21.23	22.53

* The means and standard deviations are computed from ungrouped data, not from the frequency distributions.

VII. INCONSISTENCIES, IRREGULARITIES, ERRORS

In any extensive program conducted at widely separated points, inconsistencies and irregularities of various kinds are likely to creep in; these are sometimes the source of serious inefficiency or injustice in the selection and disposition of personnel. In the present data, evidence of irregularity or error is not lacking. This section of the report presents such inconsistencies, irregularities, or errors as became apparent during analysis of the data.

A. Evidence from Punched Cards

It was not, of course, possible for us to verify the accuracy with which the Hollerith cards had been punched. Certain instances of incompleteness or error, however, were obvious. For example, a considerable number of cards (1,858, or about four per cent) were entirely blank for school-grade-completed. Table XVII below lists the types of error found in a sample of 364 cards which were rejected for some reason other than absence of school-grade-completed (in a number of cases, the cards contained more than one error each).

TABLE XVII

Types of Error Found in a Sample of Rejected Cards

Ambiguous punching of school-grade-completed (blank for one digit)	268
Unreasonable figure for school grade completed (25-95 inclusive)	45
Ambiguous punching of test score (blank for one digit)	36
Unreasonable figure for age (0-16 inclusive and 46-95 inclusive)	27
Punched 22 for location of Service School.	15
Exact or partial duplication of another card	6
Other errors	18

Under "other errors" in Table XVII are included such errors as data punched in wrong columns, double punching in one or more columns of the card, missing data, and incomplete data. Altogether, 2,184, or about five per cent, of the cards had to be discarded from all score tabulations. As the study progressed, additional cards were found containing an error or omission for individual tests; such cards were put aside during the tabulation of the test concerned, and then returned to the file for use in the tabulation of other tests. In general, a card was not discarded for error unless the error was pertinent to the data being tabulated.

B. Impossible Test Scores

Since the examinee's score on the Arithmetic test is five times the number right, each score should be a multiple of five*; similarly, on the Spelling test, each score should be a multiple of two.* A considerable number of cards were found, however, which deviated from this rule (see Tables XVIII and XIX). Such errors may have arisen either in the scoring of the papers or punching of the Hollerith cards. In our statistical analysis a card with an impossible test score was eliminated only from the particular distribution in which the defective test score would be involved.

C. Special Peculiarities of Frequency Distributions

Some noteworthy peculiarities are apparent in the distributions for certain of the tests. Specifically, in the Arithmetic test (Table XIV), the percentage of cases in the top category, 95-99, is three times as large at Great Lakes as at any other station. On the Mechanical Aptitude test (Table XIII), a heavy excess of scores is observed in the interval 0-4 for Great Lakes; while at Norfolk the excess is even greater in the interval 10-14. Another peculiarity of the Mechanical Aptitude test is the discrepancy in shape of the distributions for Great Lakes and Norfolk vs. Newport and San Diego; the two former distributions tend to be rectilinear (little tapering of frequencies at the ends of the distribution), while the latter distributions are of more nearly customary form. On the English test (Table XV), the Norfolk station is characterized by a surprisingly heavy proportion of low scores in general, and of zero scores in particular.** Charts III-V give a graphic portrayal of some

*This is true except for perfect scores of 100, which are punched as "99" in order to avoid using three columns of the Hollerith card for any one test score.

**The average score for Norfolk on the English Test is 55; this may be compared with the average of 66 for Newport, 71 for Great Lakes, and 64 for San Diego (Table XV, p. 56).

TABLE XVIII

Arithmetic Scores Not Multiples of 5

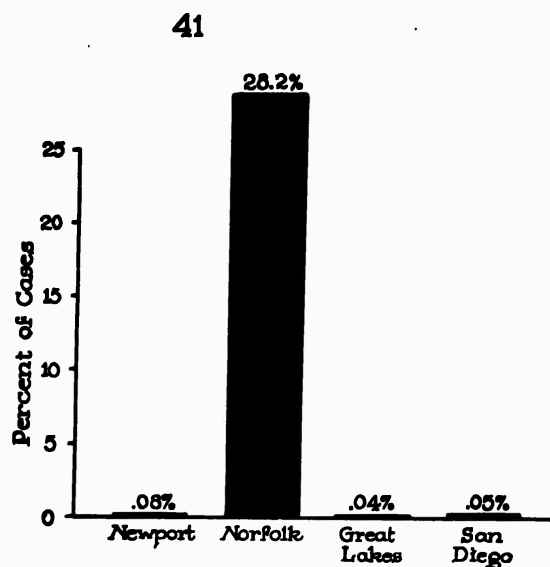
Station	Number of Cases	Scores Not Multiples of 5*	
		Number	Per cent
Newport	4864	42	.9
Norfolk	11755	258	2.2
Great Lakes	15820	176	1.1
San Diego	7808	85	1.1

TABLE XIX

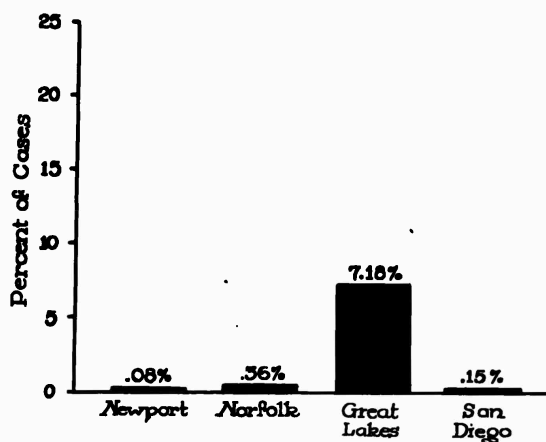
Spelling Scores Not Multiples of 2

Station	Number of Cases	Scores Not Multiples of 2*	
		Number	Per cent
Newport	4852	51	.6
Norfolk	11759	405	3.5
Great Lakes	15199	172	1.1
San Diego	7068	252	3.5

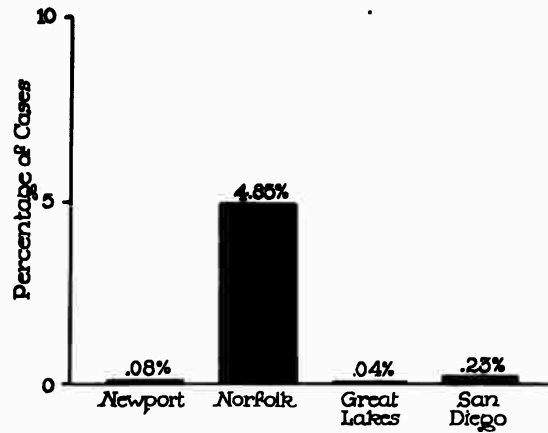
* Not including scores of 99.



**CHART III . PERCENTAGE OF CASES
REPORTED WITH SCORE OF 10 IN MECHANICAL
APTITUDE TEST AT EACH STATION**



**CHART IV PERCENTAGE OF CASES
REPORTED WITH SCORE OF ZERO IN MECHANICAL
APTITUDE TEST AT EACH STATION**



**CHART V. PERCENTAGE OF CASES
REPORTED WITH SCORE OF ZERO IN
ENGLISH TEST AT EACH STATION**

of these facts.*

Peculiarities such as those mentioned above probably do not arise merely from clerical errors in the punching of Hollerith cards. The excess of high scores in Arithmetic at Great Lakes is probably due wholly or mainly to the fact that the men at Great Lakes were given a review in arithmetical computation prior to taking the tests, whereas the men at the other stations were not. The rectilinear form of the scores for Mechanical Aptitude at Great Lakes and Norfolk evidently arises from the use of percentile norms at these stations. The causes for the other peculiarities, unfortunately, cannot be so certainly stated. One possibility that always exists is the failure of men to make sufficiently heavy marks on the answer sheets; another is the failure to make use of the special electrographic pencils; another is the possibility of copying (especially if men are seated close together); another, a mistake in timing during the administration of the tests; another, the use of an inaccurate or improper scoring-key in the test-scoring machine; another, the failure to detect improper functioning or operation of the test-scoring machine; still another, some lack of uniformity in the use of norms (especially when two different kinds of norms are available for a test, or a given set of norms is in one way or another incomplete or peculiar**). Some of these possibilities, to be sure, are not very probable. Whatever the effective causes may be, it seems certain that inconsistencies, irregularities, or errors have entered to limit the comparability or validity of scores on some of the tests.

D. Means and Standard Deviations based on Cards versus Re-Scored Papers

As was mentioned in an early part of the present report, certain highly accurate data were available from Report No. 1, based on rescored

*In Charts III-V we have plotted the percentage of cases having a score of exactly zero or exactly ten, rather than the percentage of cases in the interval 0-4 or 10-14.

**In particular, the norms for the Mechanical Aptitude test are not as complete as they should be. According to these norms, a raw score of 133 is equivalent to a percentile score of 11; but no percentile scores are given for raw scores below 133.

If low scores on the MAT were assigned a percentile rank of 10 at Norfolk and 0 at Great Lakes, this would account for part of the concentration of scores at these points for these stations. But other factors probably also play a part, since it seems unlikely that there should be so many men with such very low scores at either Norfolk or Great Lakes.

papers. The rescored papers represent only a small sample (approximately 250 cases from each of the four stations); in consequence, complete agreement between results from these papers and from the cards could hardly be expected. Nevertheless, extreme disagreement might well lead to question. Furthermore, we should expect that the differences among stations would be greater when based on small samples than when based on large, because in the case of small samples the true differences among the stations tend to be augmented by fluctuations of sampling.

As a matter of fact, however, the differences among stations are notably smaller on the basis of the small sample of rescored papers than on the basis of the 45,000 cards. Thus, the lowest and highest mean scores in GCT, from the rescored papers, are 65.51 and 71.26, respectively (range of means, 5.75); on the basis of the 45,000 cards, the corresponding figures are 65.72 and 77.02 (range, 11.30) (see Tables II and III). All the other tests, it may be emphasized, show a considerably greater difference in range of means than the GCT; the English test (to choose the most glaring case), ranges in mean from 65.45 to 70.21 on the basis of the rescored papers, and from 33.55 to 71.52 on the basis of the 45,000 cards.

The same facts are found when, instead of the means, the standard deviations in the two sets of data are considered (see again Tables II and III). Thus, in GCT, the standard deviations based on the rescored papers run from 14.62 to 17.61 (range of S.D.'s, 2.99); based on the 45,000 cards, the corresponding figures are 15.18 and 20.59 (range, 7.45). Again, the data from GCT provide only a modest estimate of the degree of discrepancy.

The irregularities mentioned in sections b and g above, for the Mechanical Aptitude and English tests, would contribute to the results that have been noted in the paragraphs above. It deserves emphasis, however, that the discrepancies we are here concerned with appear not only in the case of the MAT and English tests, but also in the case of the GCT, Arithmetic, and Spelling tests. The results from the 45,000 cards thus would appear to confirm the conclusion reached in Report No. 1, namely, that variations in procedure from one station to another constitute a problem of fundamental importance and serious magnitude.

E. Heterogeneity of Regression Coefficients

As was to be expected, and as is shown in Tables X and XI above, all of the tests show positive correlation with each other. The correlations between any two particular tests might be expected to be similar at the different stations; as a matter of fact, however, considerable differences may be observed. Thus, the correlations between MAT and GCT range from .42 (at Great Lakes) to .63 (at San Diego); between GCT and Arithmetic, from .49 (at Great Lakes) to .61 (at Norfolk); etc. A

primary question is whether the variation of correlation coefficients among the different stations is greater than might be expected from random fluctuation due to chance. If the variation exceeds that expected by chance, the next question concerns the possible sources of such true or non-chance heterogeneity.

Among the possible sources of non-chance heterogeneity in correlations are differences between so-called "alternate forms" of a test; inconsistencies or errors in the administration of tests; inconsistency in the use of norms; errors in scoring; etc. One further possible cause of heterogeneity, however, is the purely technical factor of differing range of scores at the various stations. Other things being equal, the wider the range of scores (as measured by the standard deviation), the higher the correlation coefficients will be. In order to avoid the influence of this factor, we may examine not the correlation coefficients between tests, but the regression coefficients. The regression coefficient of Y on X indicates the average rate of change of scores in Y per unit change of score in X; briefly and non-technically, it shows how closely the scores in Y reflect a change of scores in X. To pick an extreme and simple case, if the correlation between X and Y is zero, then the average Y score for men making a low score in X is the same as for men making a high score in X; the scores for Y do not at all reflect changes in the scores for X; and the regression coefficient equals zero. In such a case, the regression of X on Y -- or the sensitivity of X to changes of score in Y -- is also zero. In general, however, the regression of X on Y differs somewhat from that of Y on X (unless the standard deviations of X and Y are equal).

The regression coefficients between all the various pairs of tests (GCT and MAT, GCT and Arithmetic, etc.) are given in Table XX. The homogeneity of the regressions between MAT and GCT at the four stations -- and similarly between MAT and Arithmetic, MAT and English, etc. -- was tested by the procedure outlined in Snedecor's Statistical Methods, Chapter 12. The value of "F" for each set of regression coefficients at the four stations is given in Table XX. The "F" values in this table are to be interpreted as follows. If F equals 2.60 or less, the likelihood that the observed variation in regression coefficients would have occurred as a result of random fluctuation from a homogeneous universe is 5 per cent or more*; in other words, it is reasonable to assume that the set of regression coefficients is homogeneous. Of the twenty sets of regression coefficients in Table XX, only three meet this standard; namely, the regressions of Arithmetic on GCT ($F = 2.3$), Spelling on GCT ($F = 0.4$), and Spelling on Arithmetic ($F = 1.5$). If the value of F is between 2.60 and 3.78, the likelihood is between 1 and 5 per cent that the observed variation in the set of regression coefficients would have

*See the Table of F in G. W. Snedecor's Statistical Methods, pp. 184-187 (third edition, 1940).

TABLE XX

Regression Coefficients for All Pairs of Tests,
According to Station*
(n = 500 for Each Station)

	GCT	MAT	Arithmetic	English	Spelling
GCT					
Newport51	.81	.52	.85
Norfolk73	.86	.74	.78
Great Lakes88	1.01	.52	.87
San Diego	...	<u>.61</u>	<u>.77</u>	<u>.60</u>	<u>.82</u>
Value of F	...	<u>6.8</u>	<u>2.5</u>	<u>9.7</u>	<u>0.4</u>
MAT					
Newport	.4242	.23	.28
Norfolk	.5458	.26	.27
Great Lakes	.2026	.13	.17
San Diego	<u>.64</u>	...	<u>.67</u>	<u>.47</u>	<u>.50</u>
Value of F	<u>36.5</u>	...	<u>10.7</u>	<u>17.5</u>	<u>8.4</u>
ARITHMETIC					
Newport	.52	.2026	.39
Norfolk	.43	.4040	.45
Great Lakes	.24	.2722	.42
San Diego	<u>.40</u>	<u>.34</u>	...	<u>.56</u>	<u>.49</u>
Value of F	<u>12.6</u>	<u>5.6</u>	...	<u>11.1</u>	<u>1.5</u>
ENGLISH					
Newport	.65	.35	.8587
Norfolk	.55	.41	.6062
Great Lakes	.56	.65	1.0395
San Diego	<u>.80</u>	<u>.59</u>	<u>.90</u>	...	<u>.85</u>
Value of F	<u>11.0</u>	<u>4.8</u>	<u>9.6</u>	...	<u>8.6</u>
SPELLING					
Newport	.41	.17	.49	.54	...
Norfolk	.46	.34	.53	.49	...
Great Lakes	.32	.28	.66	.53	...
San Diego	<u>.55</u>	<u>.31</u>	<u>.61</u>	<u>.41</u>	...
Value of F	<u>9.9</u>	<u>3.1</u>	<u>2.7</u>	<u>8.3</u>	...

*The test named along the X-axis is the "dependent" variable in each test-pair. Thus, the first entry of the first row, .51, is the regression coefficient of MAT on GCT (at Newport); while the first entry of the first column, .42, is the regression coefficient of GCT on MAT (at Newport). In general, the table is best read down the columns, rather than across the rows.

occurred by chance in sampling a homogenous universe. In this case, there is some doubt as to homogeneity, but it cannot be definitely assumed that the results are heterogeneous. Two sets of regression coefficients meet this standard; namely, the regression of MAT on Spelling ($F = 3.1$) and the regression of Arithmetic on Spelling ($F = 2.7$). The remaining sets of regression coefficients -- 15 out of 20 -- are definitely heterogeneous. The values of F for these remaining sets range from 4.8 to 36.5.

It is hard to think of any "good" reason for the heterogeneity of regression coefficients at the different stations. Why should not a change in GCT score from (say) 50 to 70 be accompanied by the same average increment in Arithmetic score at all the stations? One answer is that a score of 50 in GCT (or a corresponding score in Arithmetic) does not mean the same thing at different stations.* Such a disparity might conceivably arise through the inequivalence of so-called "alternate forms" of a given test. Serious inequivalence of alternate forms, however, is very uncommon. It appears likely that some other factor or factors must be operative. What the responsible factors are, is not revealed by the present analysis; any of the irregularities already pointed out, or still others not identified, might play a part. The heterogeneity of regression coefficients indicates definitely that inconsistency of one sort or another does exist.

*Our entire discussion has assumed linear regressions throughout. Theoretically, if the regression of Y or X were curvilinear, and scores at any one station extended much above or below the scores at any other, heterogeneity of regression would be both expected and legitimate. These conditions do not obtain in the present study.

VIII. SUMMARY

The present report answers some basic questions concerning (a) the quality of recruits at each of four large naval training stations, and (b) the aptitude tests used in the classification of recruits. The basis for the present report consists in 45,559 Hollerith cards received from the Bureau of Naval Personnel. It has been found that:

1. The men at the Great Lakes training station made higher scores, on the average, than the men at Newport, San Diego, or Norfolk; on the average, the men at Norfolk made the lowest scores. Such facts deserve consideration in the establishment of service-school quotas for the various stations.

2. As measured by the standard deviation, the extent of individual differences within the various stations may, for practical purposes, be considered fairly uniform.

3. The frequency distributions of scores indicate that the General Classification, English, and Spelling tests are too easy to provide the most efficient differentiation. The Arithmetic and Mechanical Aptitude tests are of suitable average difficulty, but fail to exhibit as much concentration of scores at the middle as one would ordinarily expect.

4. The intercorrelations among the various aptitude tests range from .54 (Mechanical Aptitude x Spelling) to .61 (General Classification x Spelling). In general, the lowest correlations are found between the Mechanical Aptitude test and others (median = .45); the highest, between the General Classification test and others (median = .55). The intercorrelations tend to be highest at San Diego, and lowest at Newport (theoretically, the intercorrelations should be fairly uniform from station to station). In general, some of the intercorrelations are sufficiently high to suggest that one or more of these tests could be eliminated from the general testing program without significant loss.

5. There is a positive relation between school grade completed and scores on the aptitude tests. The highest correlation-ratio, .55, is found between school grade completed and score in the General Classification test; the lowest, .38, between school grade completed and score in the Mechanical Aptitude test. For every test, between grades six and twelve, there is on the average a definite advance of score with increase of education; from grade twelve to grade sixteen the increase tends to be smaller (nil in the case of Mechanical Aptitude test); beyond grade sixteen (college graduation), the average change is negligible. It deserves emphasis that, among men within any given educational classification, wide individual differences in test scores occur.

6. The analysis of data brought to light certain evidence of inconsistencies, irregularities, or errors. About five per cent of the Hollerith cards had to be rejected because of incompleteness (e.g., lack of school-grade-completed) or obvious error (e.g., age of recruit reported as 95 years). Some of the test scores recorded on the Hollerith cards are impossible. The frequency distributions of scores for the Mechanical Aptitude, Arithmetic, and English tests displayed noteworthy peculiarities at certain of the stations; thus, in Mechanical Aptitude, 28.2 per cent of the men at Norfolk were reported to have a score of 10; whereas less than one-half of one per cent of the men at Great Lakes were reported with this score. Such a large discrepancy is probably due, in main part, not to a real difference between the men at the different stations, but to some inconsistency or irregularity of procedure. The correlations and regression coefficients between the tests also differ excessively from station to station, suggesting that a given score cannot be depended upon to mean the same thing at different stations. These findings confirm the conclusion reached in Report No. 1 of the present project; namely, that variations in procedure from one station to another constitute a problem of fundamental importance and serious magnitude.

REEL - C

15000

A.T.I.

3 5 5 10

TITLE: Averages, Standard Deviations, and Intercorrelations of Navy Aptitude Tests

ATI- 35510

REVISION

(None)

ORIG. AGENCY NO.

Rpt No. 4

PUBLISHING AGENCY NO.

OSRD No. 1536

AUTHOR(S): (Not known)

ORIGINATING AGENCY: College Entrance Examination Board, Research and Statistics

PUBLISHED BY: Office of Scientific Research and Development, NDRC, Applied Psychology**

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
June '43	Unclass.	U.S.	Eng.	49	tables, graphs

ABSTRACT: An examination is made of the aptitude tests used in the classification of naval recruits and of the quality of the recruits at four large naval training stations. Basis for this report was compiled from the data on 43,539 Hollerith cards. Men at the Great Lakes Training station made, on the average, the highest test scores, and men at Norfolk made the lowest. The arithmetic and general aptitude tests are of suitable difficulty; but the general classification, English and spelling tests are too easy to provide the most efficient differentiation between recruits. About 5% of the Hollerith cards had to be discarded because of incompleteness or obvious error. Great differences in the test scores of men at different stations, are probably due, in the main part, not to a real difference in men, but to some inconsistency or irregularity of procedure. It is recommended that a systematic program of continuous coordination of procedures be carried out between stations.

*Laboratory, Princeton, N. J.

**Panel

DISTRIBUTION: Copies of this report obtainable from Air Documents Division; Attn: MCIDXD

DIVISION: Education and Training (38) *23(23)* **SUBJECT HEADINGS:** Aptitude tests (11381); Naval personnel -
SECTION: General (0) *4)* Aptitude tests (66099.58)

ATI SHEET NO.: R-38-0-2

Air Documents Division, Intelligence Department
Air Materiel Command

AIR TECHNICAL INDEX

Wright-Patterson Air Force Base
Dayton, Ohio

UNCLASSIFIED PER AUTHORITY OSRD LIST #9
DATED 18-21 FEBRUARY 1946