A STUDY OF INTELLIGENCE TESTING, CLASSIFICATION TESTING, AND CLERICAL APTITUDE AND MECHANICAL APTITUDE TESTING, IN A MILITARY SETTING

by

WILLIAM F. HILL

A thesis submitted in partial fulfilment of the requirements for the degree of

MASTER OF ARTS

in the
Department of

PHILOSOPHY and PSYCHOLOGY

THE UNIVERSITY OF BRITISH COLUMBIA
September 1950
A STUDY OF INTELLIGENCE TESTING, CLASSIFICATION TESTING, AND CLERICAL AND MECHANICAL APTITUDE TESTING, IN A MILITARY SETTING

by

WILLIAM F. HILL

The purpose of this study was to investigate certain psychometric procedures, and to ascertain their value in relation to the problems of selection and prediction for clerical and mechanical trades in the service. The tests selected were the Otis S-A (which were also marked for twenty minute performance as well as the standard thirty), the Wonderlic Personnel R Test, the SRA Primary Mental Abilities Test, the Detroit Mechanical Aptitude Test and the Detroit Clerical Aptitude Test. Included in the study were the marks obtained on a service-administered Classification Test - the Navy "G". The samples that were used were New Entry Trainees in the Canadian Navy who were about to take courses either as Writers (clerical trade) or Stokers (mechanical trade). The criterion used to evaluate the tests was the course marks obtained by the
Stokers and Writers on their final examination.

The tests were analyzed individually for types of distribution and amount of dispersion or variability. All the tests and subtests were correlated with the criterion to obtain validity coefficients. Similarly all the tests and subtests were correlated with the Otis, and intercorrelations were worked out for all the intelligence and classification tests. Multiple correlations of prediction were also calculated. The tests of the Primary Mental Abilities Test were intercorrelated for independence of "factors".

The validity correlations found were low but were considered to have practical significance. The lowness of the correlations was probably due to the restrictions placed on the sample by the effects of enlistment qualifications.

It was found that the twelve minute intelligence test, the Wonderlic, was apparently as good a measure of prediction as the thirty minute Otis. In the Primary Mental Abilities Test, the Number Test proved to be the best measure of prediction of any test or subtest for Stokers, and with the Reasoning Test was predictive of success in the Stokers' course. It also was the only test of the PMA which showed any possibilities for prediction with the Writers.

The Detroit Clerical Aptitude Test proved to be the best measure of all for predicting success with Writers. As for Stokers, the Mechanical Aptitude Test, while not as good as
the Clerical Aptitude for Writers, appeared to be useful if used in conjunction with an intelligence test. In fact, multiple correlations were worked out with the criterion and the Otis thirty minute, the Number test of the PMA and either the Detroit Mechanical or Clerical Aptitude Test, depending on whether the multiple was for Stokers or Writers. The coefficients were .41 and .47 for Stokers and Writers respectively.

Certain of the tests and subtests were found to be unsatisfactory on the basis that they did not distribute their scores in accordance with the normal curve and, in some instances, proved to be too difficult for the group. The Navy "G" Test was found to be unsatisfactory for the purposes of prediction, and the Space Test of the PMA was too difficult.
# CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>3</td>
</tr>
<tr>
<td>III</td>
<td>18</td>
</tr>
<tr>
<td>IV</td>
<td>20</td>
</tr>
<tr>
<td>V</td>
<td>41</td>
</tr>
<tr>
<td>VI</td>
<td>45</td>
</tr>
<tr>
<td>VII</td>
<td>62</td>
</tr>
<tr>
<td>VIII</td>
<td>68</td>
</tr>
<tr>
<td>IX</td>
<td>71</td>
</tr>
<tr>
<td>Bibliography :</td>
<td>i - xi</td>
</tr>
<tr>
<td>Appendix A :</td>
<td>Assessment Scale</td>
</tr>
<tr>
<td>Appendix B :</td>
<td>Stokers' Examination Form</td>
</tr>
<tr>
<td>Appendix C :</td>
<td>Intercorrelation of certain tests and subtests, classification tests and criterion.</td>
</tr>
<tr>
<td>Appendix D :</td>
<td>Score Distributions</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TABLE NO. | Following PAGE NO.
--- | ---
I | Chronological ages of 140 Stokers and 26 Writers shown as percentages of each sample 30
II | Last completed school grade of 140 Stokers and 21 Writers, shown as percentages of each sample 30
III | Level of former civilian occupation of 140 Stokers and 21 Writers shown as percentages of each sample 30
IV | Measures of variability of 154 Stokers and 38 Writers on criterion and tests under study 45
V | Correlation between criterion marks and test results for Stokers 53
VI | Correlations between criterion marks and test results for Writers 53
VII | Correlations between scores obtained on the Otis Thirty Minute Test and the other tests and subtests in the battery for the combined sample of Stokers and Writers 57
VIII | Correlation between the Otis S-A Thirty Minute Test and the Detroit Mechanical Aptitude Test and subtests. Sample 145 Stokers 57
IX | Correlation between the Otis S-A Thirty Minute Tests and the Detroit Clerical Aptitude Test and subtests. Sample 34 Writers 57
X | Correlation of total scores of intelligence and classification tests for 145 Stokers and 34 Writers 59
XI | Comparison of Intercorrelations obtained (a) by Thurstone, (b) Present Study 60
XII | Mean Validity Coefficients of Selection Tests in Clerical, Mechanical, and Other Branches of the Royal Navy 63

cont'd .....
<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>Description</th>
<th>PAGE No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>XIII</td>
<td>Validity Coefficients of the Army General Classification Test (U.S. Army)</td>
<td>63</td>
</tr>
<tr>
<td>XIV</td>
<td>Comparison of mean validity coefficients for the tests of the Basic Battery, U.S. Navy, Forms 1 and 2. (Correlations corrected for curtailment in ranges of talent).</td>
<td>63</td>
</tr>
<tr>
<td>XV</td>
<td>Intercorrelations among the tests and subtests of the Basic Battery, U.S. Navy, on a national sample of 500</td>
<td>63</td>
</tr>
<tr>
<td>XVI</td>
<td>Validity coefficients of unspecified intelligence tests for various types of employment</td>
<td>64</td>
</tr>
</tbody>
</table>
CHAPTER I

STATEMENT OF PROBLEM

As this study is part of a larger project involving other research workers, it is necessary to make two statements of the problem. The larger project, of which this investigation is an integral part, is under the auspices of the Defence Research Board and is concerned with the study of various approaches, psychological in nature, to the problem of selecting and classifying personnel in the Armed Forces. This involves an analysis of various techniques and tests which have been or might be applied to military personnel, and the determination of their comparative value in the military situation. In general terms the tests and techniques concerned include intelligence tests, classification tests, tests of mechanical and clerical aptitude, interest tests, trade tests and interviews.

1. Defence Research extra-mural project DRB-48 was initiated in 1947 under the direction of Prof. J.E. Morsh. The first sub-project has been reported by D. Gregory (60). The present study is one of three related studies for which preliminary work was done jointly by Shirran, Blewett and Hill. Reports will be rendered separately by Shirran on trade tests and Blewett on interest tests and the present report on intelligence and aptitude tests. The writer wishes to acknowledge the substantial assistance given this study by Defence Research Board.
The particular problem which is the basis of this study (and which is part of the total project) is concerned with certain tests of intelligence, mechanical and clerical aptitude and classification tests, and their application to military personnel. More specifically, it is a comparative study of these tests to discover their relative merits for predicting success of personnel in certain clerical and mechanical courses in the Service.
A. Mental Testing up to 1916-1917

Interest in the field of military psychology was first manifested by Plato who, in his *Republic*, cites the problem for the ideal state. The factors or characteristics which constitute the ideal soldier are outlined, and it is proposed that certain tests should be devised to discover if an individual is suited for this occupation. With the advent of the Dark Ages this problem, like many others, was left dormant, and it was not until 1905, when Zieben in Germany developed certain tests for screening defectives in the Army, that any interest was evinced in selecting satisfactory candidates for military training.

In 1908 Binet took up the problem for the French Army at the request of the Minister of War. Binet and Simon carried out some preliminary testing on soldiers, and in 1910 published an article on their progress. This article was presented to a congress of psychiatrists who apparently misunderstood the
nature of the plan and condemned the project. It was the intention of Binet and Simon to discover minimum levels for admission and thus screen those who would not be desirable for the Army. It is noteworthy that these investigators conceived that their testing would take the form of what is now called group testing. Binet died in 1911, and the type of military testing envisaged did not come into being until the advent of World War I (161).

The field of mental testing, however, was antecedent to this, and the now famous term "mental test" seems to have been first used by Cattell who, as early as 1890, published an article entitled "Mental Tests and Measurements" describing tests then actually in use in his laboratory (31). These tests, however, pertained to keenness of sight, reaction time, after images and so forth.

The birthdate of mental testing as it has now developed was 1905. It was in that year that Binet and Simon brought out their test of intelligence. The fundamental idea in their psychological method was what Binet called "a metrical scale of intelligence". The scale was composed of a series of tests arranged in increasing order of difficulty. Thirty tests composed the 1905 scale. The scale evoked much criticism which resulted in a series of revisions, the first of which appeared in 1908 bearing the significant title "The Development of Intelligence in Children." Binet brought out another revision in 1911, and Terman at Stanford completed two extensive revisions in 1915 and 1937. Terman's revision and his handbook The Measurement of Intelligence made the test acceptable throughout the
American school system (98).

The period of activity in the mental testing field up to 1916-1917 is summarised by Young (161) as follows:

(1) The large contribution which Cattell and his students made in the elaboration of statistical procedures and in the application of the tests in many directions.

(2) Binet's synthesis which was introduced to America by G. Goddard and carried into the educational system by Terman.

(3) The introduction of the intelligence quotient which was symbolic of a change from qualitative to a quantitative approach.

B. Mental Testing and World War I

Up to the time of World War I the main work in testing had been directed towards individual testing, but in a national emergency these tests, while good measuring devices, were impractical for large numbers of draftees. However, before the United States entered the war, Otis had completed a scheme for testing persons in groups with an intelligence scale made up from materials from older and well-recognized tests which were modified for group presentation. When the United States went to war, the Committee for Selection and Testing, under Yerkes, decided that Otis's plan was the key to the problem of
mass testing. The *Army Alpha* was evolved out of this work and was administered to one and a half million recruits. The *Army Beta* - a non-language test - was devised later to test illiterates (157).

The point of origin for group testing is just as definite as that for individual intelligence scales. Just as the latter stemmed from the work of Binet, so group intelligence testing as a major movement developed from the work done in the United States Army during World War I. The developments between the wars and during World War II are only comprehensible in the light of the history of development during the first world war, and certainly the work accomplished during this period gave mental testing tremendous publicity and prestige.

C. Mental Testing Between 1919 and 1939

When World War I was over there was great pressure to carry over the program to civilian needs in industry. A great many tests were patterned on the Otis and Army Alpha and, as these were not designed primarily for civilian use, they did not prove entirely suitable and some reaction against testing set in (49).

The pertinent developments between the wars were as follow:-

(1) Intelligence tests came into wide general use in schools and universities.

(2) Organization of the American College Examination
Tests which were revised annually.

(3) Setting up in England of the National Institute of Industrial Psychology (NIIP) and a similar development in the United States through the Department of Labor and the U.S. Civil Service Commission.

(4) Some test selection in various armed services.

(5) Development of student counselling programs.

(6) The emergence of the Clinical Psychologist.

(7) The development of factor analysis and the orientation it gave to mental testing.

(8) Development of aptitude testing.

It is with items 7 and 8 that the following two sections are concerned.

D. Aptitude Testing

The definition of aptitude according to Bingham (23, p.18) is, "a condition symptomatic of a person's general fitness, of which one aspect is his readiness to acquire proficiency - his general ability - and another is his readiness to develop an interest in exercising that ability". As expressed by Freeman (49, p.82) it is "the ability or collection of abilities required to perform a specified practical activity - not necessarily innate, i.e. ability to profit from training". It is this
last definition of aptitude that concerns us for this study and particularly those two fields which were pioneered by Munsterberg, namely, the clerical and mechanical. They are separately considered here:

(1) Clerical Aptitude Testing: Clerical aptitude tests have been in use for approximately thirty years and performance and work sample tests, such as typewriting tests, have been used almost from the beginning (17). The relative ease of measuring certain actual clerical skills, as well as the mental attributes obviously utilized in many clerical tasks, has been in large part responsible for the early expansion of clerical testing. Clerical testing had its beginnings in the second decade of this century. Munsterberg studied selection of clerical workers in 1916. Link, in 1919, worked along lines similar to present day practice and Thurstone's tests were in use in 1920. The depression of the early thirties initiated some widespread investigations. The Minnesota Employment Stabilization Research Institute, and also the Workers Analysis Section of U.S. Employment Service, conducted extensive researches on clerical aptitudes and aptitude testing. Similarly the NIIP in Great Britain carried out research in this field, and in both countries the findings and the tests developed were widely used during World War I.

(2) Mechanical Aptitude Testing: This field is customarily
divided under the following rubrics: mechanical performance tests and mechanical paper and pencil tests. The performance tests will not be considered here as they were not dealt with in this particular project.  

While America primarily interested itself in clerical aptitude testing, as opposed to mechanical, nevertheless considerable development also took place in the latter. Stenquist was one of the first to produce a paper and pencil test of mechanical aptitude - the Stenquist Mechanical Aptitude Tests I and II appeared in 1921. Another group paper and pencil test, the MacQuarrie Test of Mechanical Ability, was published in 1925. Two tests which were chiefly of the information type, the Detroit Mechanical Aptitudes Examination and the O'Rourke Mechanical Aptitude Test, appeared in 1929, the latter being extensively used on the Tennessee Valley Authority project. In 1930, the Minnesota Mechanical Ability Tests, under the authorship of Paterson, Elliott, Anderson, Toops and Heidbreder, made their appearance. This was a very significant battery of tests which were the result of very careful and exhaustive research. Of particular importance here was the Minnesota Paper Form Board Test which was a paper and pencil test of spatial relations (16).

1. The literature in this field has been summarised in a thesis for M.A. at the University of British Columbia by Gregory, as part of the same Defence Research Board project to which this study is related. The reader is referred to this thesis for a more extensive treatment of the whole field.
Many other acceptable tests have been designed and have served a very useful purpose in the field. Amongst these are the Purdue Mechanical Ability Test, the Science Research Associates Mechanical Aptitude Test and the Prognostic Test of Mechanical Abilities.

E. Factorial Analysis

A relatively new technique and approach to psychometrics came into prominence with the introduction of factorial analysis. This involves a mathematical manipulation of correlation matrices, applying the principle of parsimony, so that factors or groups of factors which are partialled out are interpreted as being primarily the causation of the correlations' characteristic. To state this in another way, through certain statistical techniques it is theoretically possible to determine what certain types of tests are actually measuring.

The first study in factor analysis was published by Spearman in 1904. Burt, in 1909, carried out a thorough investigation of general intelligence, using this technique, and his coefficients of relationship were interpreted to support a theory of Spearman's of a general factor pervasive in mental activity. In 1910, Brown administered a series of tests and, applying the techniques of factorial analysis to the correlations obtained, found results which were in opposition to the early findings. Since then hundreds of studies have been reported in defence of, or attacking, the general factor theory (43). Spearman's position underwent serious attack in
1938, when Thurstone published his work on factor analysis which stated that there were several independent mental factors. Thurstone proceeded to build a test purported to measure these "primary" factors and this was called the Chicago Test of Primary Mental Abilities (80). A modified form of this test was used in this present study.

The spiral omnibus type of intelligence has implicit in it the general factor of intelligence, and an interpretation of results is in terms of this postulate. Therefore, the Primary Mental Abilities Test stems from a different postulate and has, perforce, a different interpretation of test result. The discussions as to which approach is the correct one have been protracted, but some rapprochement has been achieved in the field as a result of the extensive studies carried out during the last war. A quotation from Vernon (148) gives some indication of this, "Although most if not all Admiralty and War Office Psychologists began their work with a somewhat suspicious attitude toward Spearman's views, the trend of our results pointed inescapably to a "g" plus group factor theory as preferable. It is noteworthy that American Psychologists also factorised heterogeneous Navy populations and concluded in favour of this stand. But as soon as we took selected populations such as officers, or mechanical trainees, correlations between mechanical and verbal tests tended to disappear and independent factors akin to those which Thurstone obtains in selected high school populations provided a more satisfactory picture".
The Guilford-Zimmerman Aptitude Survey (61) is one of the most recent and most exhaustive studies and they summarize the position as follows:

In the past, tests of intelligence, or clerical or mechanical ability have served well as far as they went. It is now realized however that they do not go far enough. Before World War II it had been shown by statistical analysis that human resources, as measured by tests, fall into rather separate and distinct traits to which Thurstone has given the general name of "primary abilities". The accelerated development of knowledge of aptitude factors or primary abilities during the war has confirmed this and served to advance considerably the information concerning them.

The Primary Mental Abilities Test, while having a respectable body of evidence to support its claims, was included in this test experimentally, with no intention of asserting or denying its position.

F. Military Testing, World War II

Testing in the United States Services: After World War I the importance of the proper and continued classification of manpower for many reasons was not emphasized in the United States Army. Consequently, little systematic effort was made to determine the special qualifications of enlisted men or officers, or to place them in duties for which they were peculiarly fitted. According to Davies (36), "The administration of classification tests, although prescribed by regulations, became a dead letter".

At a meeting in Washington, D.C., in 1940, the members of the Emergency Committee in Psychology of the National Research
Council proposed the mobilization of psychological knowledge pertaining to problems of human engineering in times of national crisis and defence. From this step grew the complex, elaborate and massive psychological machinery which characterized World War II.

An important milestone in the history of psychometrics was the establishment in World War I of the Group Examination Alpha. Not only was this the pioneer paper and pencil test of mental ability, but for more than 25 years it has held a clear title to the honour of being the most widely administered test. This honour now passes to the Army General Classification Test, whose upwards of nine million administrations leaves it no challenge to the title (122).

The AGCT is a test of "general learning ability" which was developed by Army personnel technicians, and it was completed before the first draftees arrived at reception stations in 1940. It has been given to every literate inductee since that time (146). It is a spiral omnibus type test, meeting the specifications of emphasizing spatial thinking and quantitative reasoning (41). The reliability by split half technique was .94 and with alternate forms, after a three week interval, .90. The AGCT correlated .79 with the Army Alpha, and .83 with the Otis, and with the A.C.E. Psychological Examination .79. According to Duncan (41), "It very effectively distinguished those of high learning ability from those of low or average ability and its main value was in selecting men for a large number of specialist training courses. The test was used for four and one half years
and over six and a half million personnel were tested".

In the U.S. Navy, psychological tests were used as early as 1912 but, prior to 1932, there was no organized testing program. In 1924, a General Classification Test was introduced for selection to Navy schools (36). Aptitude tests were later developed and used, but it was not until 1942 that the Bureau of Naval Personnel and the Applied Psychology Panel developed a series of new tests for enlisted personnel (26).

The tests making up the Navy Basic Classification Test Battery with the reliabilities of the tests is shown below:

<table>
<thead>
<tr>
<th>I. General Classification Test</th>
<th>.94</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Sentence Completion</td>
<td>.85</td>
</tr>
<tr>
<td>b. Opposites</td>
<td>.84</td>
</tr>
<tr>
<td>c. Analogies</td>
<td>.84</td>
</tr>
</tbody>
</table>

| II Reading Test               | .85 |

| III Arithmetic Test           | .83 |

| IV Mechanical Aptitude Test   | .97 |
| a. Block Counting             | .95 |
| b. Mechanical Comprehension   | .84 |
| c. Surface Development        | .98 |

| V Mechanical Knowledge Test   | .92 |
| a. Mechanical - Pictorial and Verbal | .89 |
| b. Electrical - Pictorial and Verbal | .90 |

This battery was administered to all inductees at recruiting centers, and assignment for training was decided on the basis of the test results (124).
Testing in the British Services: No large scale applications of psychology to military problems were carried out in Britain during the first World War, and the need for it was scarcely felt by the Navy or Army until the second year of World War II. Until 1941 such selection as existed in the Royal Navy was based mainly on educational examinations and on interviews. At that time a plan was instituted for classifying personnel at the recruiting centres. The testing procedure was as follows:(147)

1. A biographical questionnaire designed to bring out educational and occupational history, leisure interests and experience of leadership.

2. The Progressive Matrices Test.

3. Selected plates of a colour blindness test.

4. A short interview.

Additional testing was given later at the entry establishment on the following basis:

1. Modified Shipley Abstraction.

2. Modified Bennett Mechanical Comprehension Test.

3. Arithmetic.

4. Square Test of Spatial Judgment.

A total score was derived from this and was used as an all round index of potentiality.

As Vernon (149) points out, many of these tests were taken from American sources, but it should be noted that the emphasis was on "creative" rather than multiple choice responses.
In the Army the General Services scheme was introduced in 1941. Under this, all inductees went into the General Services Corps and there they were given the following tests: (148)

1. Progressive Matrices.
2. Arithmetic.
4. Verbal Test.
5. Physical agility.

A summation of obtained scores was used as an index of the man's potentialities. The main difference between testing programs in the American and British schemes is found in methods of intelligence testing. The Progressive Matrices Test was adopted to forestall criticism of educational bias in verbal intelligence tests. Also, more importance was attached in the British Service to the answers given to a biographical questionnaire and to an interview than to test results. In fact it was laid down that no man was to be rejected from any branch of the Service solely on the grounds of test scores (149).

Testing in the German Forces: Military psychology had a brief beginning in Germany during World War I, the first centre for testing being established in 1915. Due to the very small military establishment permitted Germany after the war, a strong effort was made to fill the ranks with the very best candidates and, therefore, psychological testing was favourably received. In 1927, the War Ministry issued a directive requiring all officer candidates to be tested, and during the next decade
testing was used routinely for the selection of all personnel (76).

The Luftwaffe program was continued during the first three years of the war, but in 1942 Reichmarshal Goring ordered the entire program to be discontinued. Following this, interviews and observations replaced assessment by tests. The German approach became one in which there was an attempt made to assess the "total personality" and the methodology was correspondingly subjective (149).
The approach to the problem of assessing certain tests of intelligence, classification, mechanical and clerical aptitude, was to obtain samples of military personnel who were undergoing - or would be undergoing - instruction in the Services in courses which were primarily clerical and mechanical in character. The assessment of the tests would necessarily be in terms of course performance.

It was necessary at the outset to lay down the following requirements: -

(1) Criteria for the selection of satisfactory samples;

(2) Criteria for the selection of tests which were to be administered;

(3) A criterion by which the efficacy of the tests could be judged.

When these requirements were met, then the selected tests would be administered to the samples of mechanical and clerical
trainees selected. The tests would be marked and scores on each test would be obtained for each trainee. The courses which the trainees were taking would yield a criterion score for each trainee.

The tests would then be examined and evaluated in terms of correlations worked out between the test results and the criterion scores. They would also be considered in the light of intercorrelations between test scores and from the standpoint of an analysis of each test's measurement of the sample. Similarly the Service-administered classification test would be analyzed. From this it would be hoped that some comparison of the efficacy and validity of the various measuring devices could be obtained.

Multiple correlations would also be calculated to ascertain if any specific combination of tests might yield higher validity coefficients.
METHODOLOGY

A. SELECTION OF TESTS

The number of tests in the field of mental testing which might be considered pertinent to this particular problem is considerable. Therefore some systematic approach was advisable, and the steps actually pursued were as follows:

(1) Test catalogues were obtained from the Science Research Association, the Psychological Corporation and the Ontario Vocational Guidance Centre.

(2) The catalogues were perused and all tests which purported to be measures of intelligence or clerical and mechanical aptitude were noted. Samples of these tests and their manuals were ordered.

(3) The tests themselves were perused and from this compendium five tests were selected and these were as follows:

i. Otis Self-Administering, Gamma;

ii. Wonderlic Personnel Test;

iii. S.R.A. Primary Mental Abilities Test;

iv. Detroit Mechanical Aptitudes Examination, Form A;

v. Detroit Clerical Aptitudes Examination, Form A.

Included in the study also were the results obtained on the Navy
G test which was administered to the trainees on enlistment.

The time available for testing would no doubt be limited, and this time would have to be shared with the other members of the research team, therefore practical considerations entered into the above selection and many tests were excluded on the time basis alone. The tests are now discussed individually and the reason for their inclusion in this study is indicated.

1. The Otis Self-Administering, Gamma: This test was selected primarily for the following reasons:

(a) It has simplicity of administration and ease of scoring.

(b) It was standardized on a group of comparable age.

(c) It is one of the oldest and most widely used intelligence tests, and has the confidence of the majority of the workers in the field.

(d) It is a reliable test claiming a reliability coefficient of .92.

(e) Most intelligence tests have patterned themselves on the Otis and, therefore, any observations based on this test could be extended to many other tests with some justification. In other words the Otis is the prototype of most intelligence tests.

(f) The test being of a spiral omnibus type allows for curtailment of time, and in fact the manual gives norms for a twenty minute version of the test.

2. Wonderlic Personnel Test: This test was selected primarily for the following reasons:

(a) It has simplicity of administration and ease of scoring.
(b) It is a well standardized test and has been favourably received and used fairly extensively in industry.

(c) It has satisfactory reliability - the author claiming reliability coefficients of .88 to .94.

(d) The test provides a curtailed administration time over the Otis S-A, being only a 12-minute test as opposed to the 30 minutes normally allowed for the Otis.

3. S.R.A. Primary Mental Abilities Test: This test was selected primarily for the following reasons:

(a) It has ease of scoring and is designed for ease in administration.

(b) According to Traxler (139) who worked out test-retest correlations on the factors, the factors are as stable as most aptitude or achievement tests, with the reliability coefficients for Verbal, Number and Spatial being over .80. Therefore the test has reasonable reliability.

(c) The test was standardized on 18,000 high school students and therefore can be considered to be well standardized. The age group and level of education of the sample bears some equivalence to the standardizing group and the particular form of the test selected claimed to be related in this respect.

(d) As this test represents a departure in psychometrics from the usual intelligence test, it was chosen to provide recency and contrast to the study and, also, because of its psychograph feature, it was thought that a profile for clerical or mechanical workers might be obtained. At least there was the possibility that the test might have some diagnostic value worth investigating.

4. Detroit Clerical and Mechanical Aptitudes Examination: These tests were selected primarily for the following reasons:
(a) They have relative simplicity of administration and ease of scoring.

(b) They have considerable respectability in the field and are rather widely used.

(c) The tests are well standardized.

(4) They are reliable tests and reliability coefficients are claimed for them of .85 and .90 for the Clerical and Mechanical scales respectively.

(5) Each of the tests have 8 subtests and these subtests cover a wide range of specific aptitudes and, in fact, many of the aptitude tests in these fields are made up exclusively of only one or two of these subtests. Therefore it was thought that in utilizing the Detroit tests one would be sampling a wider range of separate aptitude measures than with most tests in the field.

5. Navy "G" Classification Test: This test was added to the battery for the following reasons :-

(a) It made no encroachment on the time available for testing as it would have been administered in the Service beforehand.

(b) It would provide some insight into the sample as the test had already been administered and certain people had been rejected on the basis of their results.

(c) As a Service-administered test it would provide a check or control on errors introduced through administration of the foregoing tests and would supply an outside criterion for evaluation of consistency in test results.

B. SELECTION OF SAMPLES

1. The Problem of Obtaining a Sample

In obtaining a sample the practical aspects of the problem become important determiners and criteria for choice
of the sample must not only have theoretical significance but must be consistent with practical considerations. In attempting to obtain a satisfactory sample for the study the following criteria were employed:—

(a) The sample must be of a reasonable size; that is, it must be large enough to have statistical significance yet small enough to be obtainable.

(b) For the purposes of the experimental design, the sample should be obtained from personnel undergoing or shortly to undergo a training course in clerical and mechanical trades. This training preferably being in a military setting.

(c) The level of difficulty of the training course should be of a rather low order, where the instruction is fundamental and elementary rather than in an advanced or specialised field. In other words, it is implied that what is to be measured is clerical and mechanical competence in a broad sense.

(d) The course of training must yield some quantitative measure of the trainees' suitability in this pursuit or, to put it otherwise, there must be a suitable criterion derived from the course performance. The matter of criterion is expanded in sub-section D.

(e) The training centre must be reasonably close to Vancouver, B.C.
(f) There must be good liaison effected between the administrative group at the training centre as well as good rapport with the testees.

(g) The facilities for testing the trainees would have to meet a reasonable standard of comfort, quiet, ventilation and lighting. Also there would have to be a reasonable time set aside from the curriculum to administer the tests.

In translating the foregoing considerations into action many difficulties were encountered. The number of training establishments in the Province were few, and investigation was extended to Alberta. One of the few training establishments was at Chilliwack, B.C., which had a training unit for the Royal Canadian Ordnance Corps. Two trips were made there and on the last one the test battery was administered to 35 trainees. As this was apparently all the trainees they could produce, and in consideration of the facts that these were distributed throughout five trades, only a pass-fail criterion could be obtained, and liaison and testing conditions were completely inadequate, this pursuit was abandoned. The Royal Canadian Air Force Training Command at Edmonton were most cooperative, but they too only had groups of six and eight undergoing training, and therefore no further efforts were made in this direction.

In December, 1948, the Flag Officer, Pacific Coast, was personally interviewed, and he arranged an interview with the
Executive Officer of the Training School and his staff at H.M.S. Naden. From this interview it was ascertained that there were trainees in both the Writers (Clerical) and Stokers (Mechanical) schools, which more or less satisfied the foregoing seven criteria. It was further evident that this was the only sample obtainable which in any way did satisfy the criteria. Briefly the situation is presented here in terms of the seven criteria and, from this and the foregoing remarks, it can be seen why these groups were used for the purposes of this study.

(1) The sample of Stokers would apparently be ample for our purposes and as it developed 155 personnel were tested in this category. The situation for Writers did not appear so optimistic as this school was much smaller. However no feasible alternative could be found. As it turned out, the number of Writers was not an entirely satisfactory figure - the number being only 39.

(2) The sample obtained consisted of New Entries who were shortly to begin their training.

(3) The syllabus for the courses was investigated and seemed to approximate to the desired level of difficulty. This is dealt with more fully later.

(4) The method of evaluating course performance is in terms of percentiles and this is in keeping with
our experimental design. This is dealt with at length later.

(5) Esquimalt, B.C., was near enough to Vancouver feasibly to carry out the project.

(6) The liaison from the outset, throughout the whole testing period, and later when criterion marks were needed was of the highest.

(7) The facilities for testing were excellent. This is dealt with in more detail later.

2. Discussion of Sample Obtained

The sample used for this study was made up of 39 trainees enrolled in the Writer's course, and 155 in the Stoker's course. Several generalizations are herewith made concerning the constitution of these samples.

(a) The sample consists of all unmarried males
(b) They have had no previous naval experience.
(c) They have no civilian crime record.
(d) On enlistment the trainees were free from contagious disease and were of "good" health.
(e) All the students enrolled in these classes have undergone a six weeks orientation course in basic training. Some personnel are eliminated at this stage, usually on the basis of response to discipline.
At enlistment the trainees are given the Navy "G" classification test, and on the basis of this test persons scoring below 36 are not accepted.

It can be seen from the foregoing that the results of these six generalizations would be to affect the sample in a way which would make it more homogeneous and something less than an unselected group. While this is undeniably an undesirable feature in the sample, it is inherent in military sampling.

The sample was analyzed to discover if any particular strata of society or any particular class of person was attracted to these particular trades in the Navy during peace-time. The data for this analysis was obtained from a questionnaire designed by a fellow research worker on this project, Blewett, and it was administered to the group in the test administration sequence. The analysis deals with three aspects of the trainees:-

1. The age;
2. the formal educational level;
3. the level of former employment.

The analysis of the age and educational level is self-evident, but some explanation of the analysis of former employment is in order. In dealing with the wide variety of jobs which the trainees formerly held, some systematic grouping was necessary, and a loose socio-economic classification was utilized. Six different classifications were used as follows:-
Class I: White collar - store clerk, bookkeeper, teller, checker, etc.

Class II: Skilled workers - locomotive engineers, foremen, etc.

Class III: Semi-skilled and unskilled workers: construction, janitor, barber, etc.

Class IV: Semi-professional - salesmen, owners of small businesses, etc.

Class V: Farmers - farm hands and owners of small farms.

Class VI: Students - no work history other than casual.

(1) **Age:** From a study of Table 1 it can be seen that 78 per cent. of the Stoker sample are distributed between the ages of 18 and 20, and 63 per cent. of the Writers fall within this range. Further to this, it can be seen that 96 per cent. of the Stokers are included in the age range of 17 to 21, and 88 per cent. of the Writers are included in the age range 18 to 22. From this it can be seen that both samples are homogeneous for age, in comparison with what might be expected for a wartime induction group.

(2) **Formal educational level:** In analyzing Table 2, we find that 94 per cent. of the Stokers fall between Grades VIII and X. There is an upward shift in educational level for the Writers, but 95 per cent. of this sample is embraced in the spread between Grades IX and XIII. The spread here in the sample is also quite narrow.

(3) **Level of former employment:** Table 3 provides the
data from which the following observations are made. The sample of Stokers were primarily engaged before enlistment in unskilled and semi-skilled occupations, 76 per cent. of the sample falling in this category. As for the Writers, they either enlisted straight from school (30 per cent.) or worked in white collar jobs (55 per cent.) and only 15 per cent. had any other background. Therefore it can be stated for the samples of Stokers and Writers that there is within each group considerable similarity of pre-enlistment background.

Summarizing the foregoing, we find that the physical organisation and procedures of the Navy have made some curtailments of the normal unselected sample, and that, over and beyond this, there is evidence of external factors which seem to operate in a manner which selects somewhat similar people, similar that is in terms of background, experience and education.

C. TEST ADMINISTRATION

The test battery was first administered to a group of trainees at the Army Training Centre at Chilliwack, B.C., but as was previously mentioned these trainees were not used in the final sample. However, this enterprise, while in the main abortive, did perform a useful function in test administration. That function was to provide a trial run for the battery, and during the trial run many difficulties were encountered which were eliminated prior to the administration of the sample used in this study. Thus, the total length of time required for
### TABLE I

Chronological ages of 140 Stokers and 26 Writers, shown as percentages of each sample.

<table>
<thead>
<tr>
<th>Age</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stokers</td>
<td>8</td>
<td>34</td>
<td>28</td>
<td>16</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>Writers</td>
<td>4</td>
<td>35</td>
<td>13</td>
<td>20</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

### TABLE II

Last completed school grade of 140 Stokers and 21 Writers, shown as percentages of each sample.

<table>
<thead>
<tr>
<th>Grades</th>
<th>VII</th>
<th>VIII</th>
<th>IX</th>
<th>X</th>
<th>XI</th>
<th>XII</th>
<th>Beyond</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stokers</td>
<td>2</td>
<td>38</td>
<td>36</td>
<td>18</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>Writers</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>32</td>
<td>46</td>
<td>10</td>
<td>2</td>
<td>100</td>
</tr>
</tbody>
</table>

### TABLE III

Level of former civilian occupation of 140 Stokers and 21 Writers, shown as percentages of each sample.

<table>
<thead>
<tr>
<th>Employment Level</th>
<th>White Collar</th>
<th>Semi &amp; Skilled</th>
<th>Unskilled</th>
<th>Semi-Prof.</th>
<th>Farmer</th>
<th>Stud.</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stokers</td>
<td>7</td>
<td>3</td>
<td>76</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>Writers</td>
<td>55</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>100</td>
</tr>
</tbody>
</table>
proper administration, the optimum interval for break periods and the best sequence of tests in terms of sustained motivation, were all determined in advance. Not only was it possible, on the basis of the Chilliwack experience, to plan the best arrangement for testing, and not only was actual practice acquired in standardized administration, but equally important was the acquisition of finesse in dealing with a group of military personnel undergoing a relatively unfamiliar task under the direction of a civilian.

Three trips to Esquimalt were required to complete all the testing and, in each instance, the groups tested were given the tests in the same order, and with the same arrangement for break periods. An entire day was required to administer all the tests as there were the tests of the other members of the research team as well as those pertinent to this study.

The sequence of tests used was as follows:

**Morning**

Vocational Questionnaire  
Otis S-A Intelligence Test  
Lee-Thorpe Interest Test  
Wonderlic Personnel Test

**Afternoon**

Detroit Mechanical or Clerical Aptitude Test  
Kuder Preference Record  
S.R.A. Primary Mental Abilities Test
The standardized administration as outlined in the manual was adhered to in the administration. One additional instruction was introduced to discover the point of progress of testees at 20 minutes in the Otis. The testees were informed at the outset that the following instructions would be given at the end of 20 minutes: "Twenty minutes! Circle the question you are now doing and continue". This was felt to be a minimal distraction and would not greatly affect the scores on the 30 minute scale.

The accommodation and facilities for testing were entirely adequate. Each man had a separate desk with sufficient leg space to ensure he was not cramped or otherwise handicapped. Both the lighting and ventilation were satisfactory. As these facilities were designed and used for classroom purposes the general layout was such that cheating or copying were difficult to accomplish. The entire administration and marking of the tests was at least as objective, consistent and controlled as would be found in any military testing program and probably considerably superior.

D. SELECTION OF CRITERION

The problem of choosing a criterion or criteria is crucial to the ultimate significance of a study and yet, in the final analysis, there is no absolute criterion which is without some theoretical weakness. This is a common problem in test construction, particularly in mental testing, and many of the
difficulties are put forth by Mursell (98). He states that one method of test validation is to correlate the test with another well known test, usually the Binet. If a high correlation is obtained, then this means that the new test is measuring to a large extent what the Binet was already measuring. Stated in another way, a test which is theoretically superior to the Binet would tend to correlate lowly because of its superiority. Another criterion employed is the degree of correlation with examination marks. However, it was the unreliability of examination marks in the first place that led in part, at least, to the rise of mental testing in educational psychology. Similarly for correlations with a teacher's estimates of the students.

In the industrial field an analogous situation exists for validity criteria. These criteria have been categorized by Davies (35) and are included here:

(1) Objective records of individual performance.
(2) Difference between groups of known characteristics.
(3) Results of examinations.
(4) Gradings or assessments.
(5) Level of aspiration.

As indices of vocational success, each one of these has relative merit, but not doubt is not without some theoretical defect. Thorndike earlier used three main criteria for vocational success, namely: earnings, job status, subject's interest and and satisfaction.
Bray (26) suggests somewhat similar types of criterion, which could be applied in military situations, such as:

1. Performance on course examination.
2. Rates of pay.
3. Rank.
4. Ratings on the job by superiors.
5. Rate of promotion.

As Rodger (114) states, "entirely satisfactory criteria to judge the values of procedures is hard to find"; and to quote from OSRD #6110, quoted in Davis (36, p. 46) "the various tests and procedures used in classification of naval personnel should be validated against criteria which are as objective and realistic as possible". In so far as this study is concerned, the former remark, by Rodger, is appropriate. Of the various criteria suggested by Bray and quoted above, most of them could be eliminated for this study on either practical considerations of time, or on theoretical objections, or both. In respect to the second statement, the criterion used - course marks - has a measure of objectivity, as we will point out later, and it is also realistic. It is realistic in the sense that the trainee in either the Stoker or Writer course must pass the examinations to become a Stoker or Writer in the field.

It should be pointed out that while criteria of this type are realistic, so far as they go, they still are not optimal. In the first place, the grades tend to overemphasize "book
learning" or "theory" to the sacrifice of actual ability to perform and, further, the grades or marks do not indicate in what manner or area the individual was unsatisfactory, and, finally, personality factors may enter in unknown degrees in determining the trainees' final grade.

The criterion of course marks was selected in the final analysis, because it was judged to be the best measure of validity available within the practical limitations of the study, and to quote Davies (35), "In real life situations you have to take the criteria which you can get despite its limitations".

As course marks have been widely used in the field of military psychology, it is inferred that it is a legitimate and justifiable procedure. However, it is possible that in any particular situation course marks will be an unsatisfactory criterion of validity. Therefore it is necessary to investigate the structure of the criterion used in this particular piece of research. This was done for both the Stokers and Writers under the following headings: -

(1) Nature of the Curriculum.
(2) Nature of the Examination
(3) Nature of the Assessment Marking

It might be pointed out that, since part of the study is concerned with measures of clerical and mechanical aptitude, it would be necessary to determine whether clerical or
or mechanical functions are involved in the subject matter of the respective courses, and whether the examinations are directly related to such subject matter, before any valid conclusions could be drawn regarding the efficacy of these tests as measures of clerical or mechanical aptitude.

STOKERS

1. Nature of the Curriculum

This is a six weeks course of fairly intensive classroom instruction supplemented by trips to engine rooms aboard ships and the showing of films. There is also a set of comprehensive mimeographed notes supplied to each man covering the whole course and consisting of 140 foolscap pages. The reference text for the course is Chant's *Elementary Physics*. A brief survey of the topics covered by the course follows:

**Steam**: physical principles involved; production and control; construction of boilers.

**Boiler Room Machinery**: Function and construction of various parts and their inter-related functioning.

**Oil Fuel Systems**: Theory of combustion; construction - maintenance duties.

**Pumps**: Theory of hydraulics; duties of maintenance.

**Construction and Organisation of Cruiser**: Main features; safety devices; bulkhead design; fire-fighting equipment etc.

**Engine Room Machinery**: History of its importance in various battles.

**General Organization of Major Equipment**: Function and components and inter-relationship of the 17 major items of machinery.

**Propulsion**: Gearing, turbines, lubrication and other duties and principles of steering.
Generators; exhaust systems, internal combustion engines.

Air Compressors: theory and component parts.

It should be noted from above that the underlying physical principle is explained, the historical importance is indicated from results of malfunction in certain naval engagements, and that the duties are clearly detailed. By the use of diagrams, mock ups and visits to ships, the function of the equipment is clearly outlined.

This then is the curriculum of study for the Stokers and it can be seen that it is primarily of a mechanical nature.

2. Nature of the Examination

A copy of the final examination is to be found in Appendix B, and from a perusal of this sample examination it can be seen that it is testing a knowledge of mechanical matters. The criterion mark is solely the mark achieved by the trainee on this final examination.

3. Nature of the Assessment Marking

The criterion is made up exclusively from the marks obtained on the final examination, and therefore can be considered to be objective. It should be pointed out that the sample of Stokers is made up of three groups, and therefore different instructors and examinations are involved in the composite sample which grouped them all together for purposes of correlation. The Officer Commanding Stoker Training was
interviewed concerning this and it was felt that he was aware of the possibility of error being introduced by inconsistencies of instruction and examination. However, he has striven to prevent this by comparing performance over time, from group to group, against various instructors and various alternate forms of examination. The examinations are of apparent equality in level of difficulty and the instructors work together and are aware of the standards and teach by the curriculum. The instructors mark each other's class papers, using a key, and this tends to insure uniformity. Conferences are held by the Officer Commanding where marks are reviewed by the staff as a whole, and there seem to be no anomalies from group to group in terms of presentation, examination or scoring. Therefore it was felt that these groups could be amalgamated with some justification.

WRITERS

1. Nature of the Curriculum

The course is comprised of various subjects with different instructors for each subject. The subjects are as follows:

**Typewriting:** Twenty weeks instruction. Standard commercial school pedagogy used.

**Shorthand:** Twenty weeks instruction. Pitman's method taught at time of testing and manual used found in most commercial schools.

**Bookkeeping:** Eight weeks instruction. Elementary bookkeeping practice where such things as debits and credits, trial balances and posting are taught. Text for the course was *The Canadian Modified Accountancy*.  
Administration: Eight weeks instruction. A course acquainting the trainee with the organisation of administration and forms to be filled out and standard administrative procedures to follow.

2. Nature of the Examination

There were separate examinations for each of the courses:—

Typewriting: An examination for speed and accuracy.

Shorthand: An examination for speed and accuracy

Bookkeeping: A final examination with problems involving the keeping of books and posting ledgers and obtaining trial balances and so forth.

Administration: A final examination based on questions of procedure and also questions requiring the correct use of various military forms.

3. Nature of the Criterion Score

The final standing of the trainees was derived from a composite of separate test scores, the maximum mark being 500 and the criterion mark being expressed in percentage. The final mark was composed of a maximum of 100 marks for typewriting, 100 marks for shorthand or bookkeeping, and 200 marks for administration. Besides this there were 100 marks given for an assessment of the man.

There are two theoretical objections that could be raised against the use of these course marks as a validity criterion. (1) The amalgamation into one group of those who took shorthand as opposed to those who took bookkeeping.
(2) The assessment of the individual introduced a subjective element which is also not directly related to competency in the clerical field.

As to the first, it must be admitted that even the amalgamated group was statistically too small and certainly a further split would preclude any study of clerical trainees. Further to this, the final aggregate mark is still the Navy's evaluation of the man as a potential Writer. In respect to the second objection, it should be pointed out that an attempt was made to make the rating objective, and the derived score was obtained through the use of an Assessment Scale (Appendix A) which was filled out by the instructor with the officer-in-charge. This assessment was made in terms of whether the individual had the attributes which were thought to be related to being a satisfactory Writer.

The final justification for using these marks for criterion was that there were no other available criteria within the practical limitations of the study.
CHAPTER V

METHODOLOGY OF STATISTICAL ANALYSIS

Each test along with its subtests was scored for all the Writers and Stokers who were tested. Scores were also available for these groups on their Navy-administered "G" test. Besides these results, there were the marks obtained by the individuals on the Stoker and Writer courses respectively. This resulted in approximately 3400 separate scores. These were first organized by tabulating the test marks for each man in order of tests and criterion, so that opposite each man's name there were all the test and criterion scores achieved.

The possibilities for working and reworking this data were almost infinite. Item analysis, intercorrelations, analysis by multiple correlations, factorial analysis, covariance techniques and so forth, could be employed and probably information of importance could be derived. However, some restraint must be imposed by the nature of the problem, and the attending practical considerations. Thus a selection
had to be made from all the possible statistical methods of analysis and this was done in the light of practical considerations and the experimental design. The analysis was conducted in the following manner.

A. ANALYSIS OF TESTS AND SUBTESTS IN TERMS OF DISTRIBUTION AND VARIABILITY

i. Stokers

ii. Writers

Each test was examined individually as well as the criterion scores to determine if they were legitimate measuring devices. Attention was specifically directed towards the degree of variability of the scores and the nature or characteristic of the distribution of the same scores.

B. ANALYSIS OF CORRELATIONS OF TEST SCORES WITH CRITERION

i. Stokers

ii. Writers

Each test and subtest was correlated with the criterion using Pearson's Product Moment Technique, and the coefficients of validity thus obtained were considered to be relative measures of validity and indicators of the particular test or subtest's relative value for predicting performance on the Stoker's and Writer's courses.
C. ANALYSIS OF CORRELATION OF TEST SCORES WITH THE OTIS S-A

The Stokers and Writers were grouped together and the test scores for all the tests were correlated with the Otis S-A 30-minute version. This was done to obtain some insight into the extent to which a test or subtest tended to measure the same area as the Otis, and also on the basis of this the analysis should supply indications for selecting test correlations for a multiple correlation.

D. ANALYSIS OF INTERCORRELATIONS OF TOTAL SCORES OF INTELLIGENCE AND CLASSIFICATION TESTS

i. Stokers

ii. Writers

All the intelligence tests and the classification test were intercorrelated and the results are analyzed to see if any marked superiority or anomalies pertained from measure to measure.

E. ANALYSIS OF INTERCORRELATION OF TESTS OF THE PRIMARY MENTAL ABILITIES TEST

All the test scores for the Primary Mental Abilities Test were intercorrelated and in this case the Stokers and Writers were grouped together. The purpose of this was to determine to what extent the Primary Abilities Tests were independent.
F. ANALYSIS OF MULTIPLE CORRELATIONS

i. Stokers

ii. Writers

Multiple correlations for the Stokers and Writers were computed to discover if certain combinations of tests significantly raised the validity coefficients.

G. ANALYSIS OF SCATTERGRAMS

Scattergrams were constructed for the correlations, and these were analyzed to discover the nature of the relationship and the effects of cut-offs.
CHAPTER VI

ANALYSIS OF EXPERIMENTAL RESULTS

A. DISTRIBUTION AND VARIABILITY OF TESTS AND SUBTESTS

The data from which this analysis is made are found in Table 4 and Appendix D, Figures 1-40.

1. Stokers

**Criterion - Fig. 1:** The score results follow a distribution which is essentially that of the normal curve. From an inspection of the coefficients of variability it can be seen that the criterion has the least variability of all the measures. This narrow dispersion of scores indicates the homogeneity of the group.

**Otis S-A Thirty Minute Test - Fig. 2:** Distribution is essentially normal and the amount of variability is satisfactory for an intelligence test.

**Otis Twenty Minute Version - Fig. 3:** Distribution is normal and the variability satisfactory.
<table>
<thead>
<tr>
<th>Tests</th>
<th>Writers</th>
<th>Stokers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stan. Dev.</td>
<td>Coeff. of V.</td>
</tr>
<tr>
<td>Criterion</td>
<td>78.59</td>
<td>6.32</td>
</tr>
<tr>
<td>Otis 30 min.</td>
<td>47.67</td>
<td>10.29</td>
</tr>
<tr>
<td>Otis 20 min.</td>
<td>39.65</td>
<td>8.11</td>
</tr>
<tr>
<td>Wonderlic</td>
<td>24.50</td>
<td>5.56</td>
</tr>
<tr>
<td>PMA Total</td>
<td>150.79</td>
<td>30.18</td>
</tr>
<tr>
<td>PMA #1</td>
<td>31.62</td>
<td>4.49</td>
</tr>
<tr>
<td>PMA #2</td>
<td>20.79</td>
<td>12.66</td>
</tr>
<tr>
<td>PMA #3</td>
<td>16.23</td>
<td>5.04</td>
</tr>
<tr>
<td>PMA #4</td>
<td>32.61</td>
<td>12.55</td>
</tr>
<tr>
<td>PMA #5</td>
<td>50.02</td>
<td>10.24</td>
</tr>
<tr>
<td>Det. Apt. #1</td>
<td>29.64</td>
<td>3.11</td>
</tr>
<tr>
<td>Det. Apt. #2</td>
<td>20.05</td>
<td>2.60</td>
</tr>
<tr>
<td>Det. Apt. #3</td>
<td>32.56</td>
<td>7.35</td>
</tr>
<tr>
<td>Det. Apt. #4</td>
<td>33.79</td>
<td>5.81</td>
</tr>
<tr>
<td>Det. Apt. #5</td>
<td>19.97</td>
<td>3.60</td>
</tr>
<tr>
<td>Det. Apt. #6</td>
<td>22.33</td>
<td>8.48</td>
</tr>
<tr>
<td>Det. Apt. #7</td>
<td>26.44</td>
<td>4.96</td>
</tr>
<tr>
<td>Det. Apt. #8</td>
<td>23.38</td>
<td>12.18</td>
</tr>
<tr>
<td>Navy G Test</td>
<td>52.00</td>
<td>10.60</td>
</tr>
</tbody>
</table>
Wonderlic - Fig. 4: Distribution essentially normal and the variability satisfactory.

P.M.A. Total Score - Fig. 5: The distribution follows in effect the expression of the normal curve, and the dispersion is in keeping with that of the other intelligence tests in the battery.

P.M.A. # 1, Verbal Meaning - Fig. 6: Distribution is essentially normal and somewhat negatively skewed. This test shows more variability than most.

P.M.A. # 2, Space - Fig. 7: The distribution here indicates that the test was too difficult for a large number of testees, and consequently they are not measured by the test. This is a very undesirable feature. The amount of variability is also high in relation to the other tests.

P.M.A. # 3, Reasoning - Fig. 8: The distribution is fairly normal and the coefficient of variability is higher than that for most of the tests.

P.M.A. # 4, Number - Fig. 9: Distribution tends towards the normal. The variability of the group is relatively high.

P.M.A. # 5, Verbal Facility - Fig. 10: Distribution is a reflection of the normal curve and is somewhat negatively skewed, indicating that the test is somewhat too easy for the group. The variability is relatively normal for the test.

Detroit Mechanical Aptitude Total - Fig. 11: The distrib-
ution of scores is closely congruent with the normal curve, but the amount of dispersion is somewhat low.

**Detroit Mechanical Aptitude Test #1 - Fig.12:** The distribution follows the normal curve but is somewhat leptokurtic and this is borne out by a low coefficient of variation.

**Detroit Mechanical Aptitude Test #2 - Fig.13:** The distribution is essentially normal and the amount of dispersion satisfactory.

**Detroit Mechanical Aptitude Test #3 - Fig.14:** The distribution is normal and dispersion acceptable.

**Detroit Mechanical Aptitude Test #4 - Fig.15:** The distribution is somewhat normal except that a disproportionate number found the test too difficult to make a score on it. The dispersion is satisfactory.

**Detroit Mechanical Aptitude Test #5 - Fig.16:** The distribution here is entirely unsatisfactory and bears no relation to the normal curve. The distribution is somewhat rectangular and a large loading of people found the test too difficult.

**Detroit Mechanical Aptitude Test #6 - Fig.17:** Essentially normal distribution and the spread of the scores is quite satisfactory.

**Detroit Mechanical Aptitude Test #7 - Fig.18:** This test is too easy for the group and a preponderant number obtained a
perfect score. However the rest are spread out evenly and the
test could be useful for eliminating personnel at a cut-off
point. The variability is fairly high.

Detroit Mechanical Aptitude Test # 8 - Fig. 19: Essentially
a normal curve and the dispersion of the group is satisfactory.

Navy "G" Test - Fig. 20: The test was used for acceptance
into the Navy at recruitment and the critical score was 36.
From the distribution here we can see that most of the Stokers
fall at the cut-off point and taper off to the top end normally.
The dispersion of the group is small and the curtailment of
range by the previous use of the test is graphically illustrat-
ed here.

Summary

(1) The criterion while satisfactorily distributed is too
restricted in the dispersion of scores.

(2) The intelligence tests are adequate for distribution
and variability.

(3) Test # 2 - Space - of the PMA is unsatisfactory, but the
other tests have reasonable utility.

(4) The Detroit Mechanical Aptitude Test Total Score is a
legitimate measuring device for this sample.

(5) Most of the subtests of the Detroit are satisfactory
with the exception of # 4 which has a tendency to be too
difficult and Test # 5 which is undoubtedly unsatisfactory.
Test # 7, while being too easy, could still be used.

(6) The Navy "G" Test used in this context, with the curtailment effects from its previous administration being in evidence, is unsatisfactory.

ii. Writers

Criterion - Fig.21: The distribution of scores resembles the normal curve, but is somewhat leptokurtic, and this is reflected in the dispersal of scores which is low and, in fact, the coefficient of variation is the lowest for this measure of any for either Writers or Stokers.

Otis S-A Thirty Minute Test - Fig.22: The normal curve is approximated with a tendency towards a platykurtic type. The dispersion of scores is satisfactory.

Otis Twenty Minute Version - Fig.23: Distribution of scores is normal and the variability of the group is satisfactory.

Wonderlic - Fig.24: There is some similarity of distribution with the normal curve with negative skewness and satisfactory dispersion.

PMA Test Total - Fig.25: The trend is toward the normal curve of distribution but with some positive skewness. The spread of the scores is satisfactory.
PMA Test # 1 - Verbal Meaning - Fig. 26: There is an approximation of the normal curve but with a rather small dispersion for the group.

PMA Test # 2 - Space - Fig. 27: This test is too difficult for the group and many of the testees were unable to score on it, thus the group were not dispersed satisfactorily.

PMA Test # 3 - Reasoning - Fig. 28: There is the suggestion of bimodality here but this might change with the addition of a few cases into the normal curve. The dispersion of the group is relatively high as a consequence.

PMA Test # 4 - Number - Fig. 29: The curve here is essentially normal and the spread of the scores is somewhat high.

PMA Test # 5 - Verbal Facility - Fig. 30: A normal distribution of scores with an average coefficient of variation.

Detroit Clerical Aptitude Test Total Score - Fig. 31: It approximates the normal curve but has a small dispersion of scores.

Detroit Aptitude Test # 1 - Fig. 32: There is a tendency toward the normal curve, but it is positively skewed and with a small dispersal of scores.

Detroit Aptitude Test # 2 - Fig. 33: Essentially a normal curve but the dispersal of scores is also quite low.
Detroit Aptitude Test # 3 - Fig. 34: There is little resemblance to the normal curve and some evidence of bimodality with the scores spread unevenly. The dispersal of scores is consequently higher.

Detroit Clerical Aptitude Test # 4 - Fig. 35: The normal curve is evidenced here and the spread is somewhat higher but still a little low.

Detroit Clerical Aptitude Test # 5 - Fig. 36: An expression of the normal curve with a dispersal of scores similar to # 4.

Detroit Clerical Aptitude Test # 6 - Fig. 37: Approximating the normal curve but with a larger variability in score spread.

Detroit Clerical Aptitude Test # 7 - Fig. 38: The distribution is normal but the spread is small.

Detroit Clerical Aptitude Test # 8 - Fig. 39: There were four individuals who could not score on the test which distorted the otherwise normal curve, and the spread of the scores was very high - in fact the highest for any test for either group.

Navy "G" Test - Fig. 40: The distribution conforms to no particular pattern and the scores are spread along the baseline so that it appears that the ability to do the test is spread diversely throughout the Writers, as opposed to the tapering off of the Stokers.
Summary

(1) The criterion, while evidencing a satisfactory distribution of the scores, is too restricted in its score spread.

(2) The intelligence tests on the whole are satisfactory measures of the group.

(3) Test #2 - Space - of the PMA is no better for the measurement of Writers than Stokers and is too difficult for the group. The other tests are relatively useful measuring tools.

(4) The Detroit Clerical Aptitude Test Total Score is a fairly good measurement of the group but has somewhat limited dispersion of the scores.

(5) Most of the subtests of the Detroit Clerical Aptitude Test are satisfactory with only Test #3 being somewhat bimodal, but with a larger sample the normal curve might emerge distinctly.

(6) The Navy "G" Test is not a satisfactory measure of this sample.

General Observations

While certain tests have been indicated as being unsatisfactory or at least not entirely desirable (Stokers - PMA #2; Detroit Aptitude #4, #5 and #7; Navy "G" : Writers - PMA #2; Detroit Aptitude #3; Navy "G") there are several tests which
while satisfactory for the sample might present difficulties with larger groups. Where the top or bottom scores already are represented at the maximum or minimum score intervals, it would indicate that the test has perhaps not enough "top" or "bottom" to discriminate certain high or low scoring groups where large numbers are being tested. From Appendix C it can be seen that for the Stokers this applied to PMA #1 and #2 and #4, to Detroit Mechanical Aptitude #1, #5, #6, #7 and #8; and for the Writers to PMA #1, #2, #4, and #5 and Detroit Clerical Aptitude #8.

B. ANALYSIS OF CORRELATIONS OF TEST SCORES WITH CRITERION

One hundred and eight intercorrelations were calculated from the raw scores between criterion, tests and subtests, using Pearson's Product Moment Technique (52), and these are represented in a master chart in Appendix C. From these correlations certain tables have been abstracted to provide a more meaningful analysis. Tables 5 and 6 are an abstraction of all the correlations of the tests and subtests with the criterion for Stokers and Writers respectively.

i. Stokers

**Intelligence and Classification Tests:** The Thirty Minute Otis and the Wonderlic correlated with the criterion .27 and .25 respectively which was significant to the .01 level of confidence. The Otis Twenty Minute Version, the total score on the PMA and the Navy "G" were all .16 correlations which were sig-
### TABLE V
Correlation between criterion marks and test results for Stokers.

<table>
<thead>
<tr>
<th>Criterion and Otis Thirty Minute</th>
<th>Coeff. of Correl.</th>
<th>Sig. at level</th>
<th>Number of Testees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otis Twenty Minute</td>
<td>.27</td>
<td>.01</td>
<td>144</td>
</tr>
<tr>
<td>Wonderlic</td>
<td>.16</td>
<td>.05</td>
<td>90</td>
</tr>
<tr>
<td>PMA Total Score</td>
<td>.25</td>
<td>.01</td>
<td>143</td>
</tr>
<tr>
<td>PMA #1, Verbal Meaning</td>
<td>.16</td>
<td>.05</td>
<td>140</td>
</tr>
<tr>
<td>PMA #2, Space</td>
<td>.07</td>
<td>Not</td>
<td>140</td>
</tr>
<tr>
<td>PMA #3, Reasoning</td>
<td>.21</td>
<td>.01</td>
<td>140</td>
</tr>
<tr>
<td>PMA #4, Number</td>
<td>.34</td>
<td>.01</td>
<td>140</td>
</tr>
<tr>
<td>PMA #5, Word facility</td>
<td>.02</td>
<td>Not</td>
<td>140</td>
</tr>
<tr>
<td>Detroit Mech. Total Score</td>
<td>.21</td>
<td>.01</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #1, Tool Inf.</td>
<td>.12</td>
<td>Not</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #2, X in circle</td>
<td>.01</td>
<td>Not</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #3, Est. of size</td>
<td>.16</td>
<td>.05</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #4, Arithmetic</td>
<td>.22</td>
<td>.01</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #5, Space org.</td>
<td>.10</td>
<td>Not</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #6, Mech. Inf.</td>
<td>.15</td>
<td>Not</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #7, Direct. Pull.</td>
<td>.12</td>
<td>Not</td>
<td>145</td>
</tr>
<tr>
<td>Detroit Mech. #8, Digit-alpha.</td>
<td>-.01</td>
<td>Not</td>
<td>145</td>
</tr>
<tr>
<td>Navy G. Classification Test</td>
<td>.16</td>
<td>.05</td>
<td>145</td>
</tr>
</tbody>
</table>
TABLE VI

Correlations between criterion marks and test results for Writers.

<table>
<thead>
<tr>
<th>Criterion and Test</th>
<th>Coeff. of Correl.</th>
<th>Signif. at Level</th>
<th>Number of Testees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thirty Minute Otis</td>
<td>.29</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>Twenty Minute Otis</td>
<td>.35</td>
<td>.05</td>
<td>29</td>
</tr>
<tr>
<td>Wonderlic</td>
<td>.25</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>PMA Total Score</td>
<td>.27</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>PMA #1, Verbal Meaning</td>
<td>.10</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>PMA #2, Space</td>
<td>.13</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>PMA #3, Reasoning</td>
<td>.10</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>PMA #4, Number</td>
<td>.27</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>PMA #5, Word fluency</td>
<td>-.02</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical Total Score</td>
<td>.44</td>
<td>.05</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #1, Handwriting</td>
<td>.12</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #2, Item Checking</td>
<td>.44</td>
<td>.05</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #3, Arithmetic</td>
<td>.32</td>
<td>.05</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #4, X's in circle</td>
<td>.10</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #5, Clerical inf.</td>
<td>.11</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #6, Space organ.</td>
<td>.14</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #7, Digit-Alpha.</td>
<td>.56</td>
<td>.05</td>
<td>34</td>
</tr>
<tr>
<td>Detroit Clerical #8, Filing</td>
<td>.01</td>
<td>Not</td>
<td>34</td>
</tr>
<tr>
<td>Navy G Classification Test</td>
<td>.19</td>
<td>Not</td>
<td>33</td>
</tr>
</tbody>
</table>
significant to the .05 level.

**Primary Mental Abilities Tests**: The numbers test, #4, correlated the highest of any of the tests or subtests administered to the Stokers, correlating .34. Reasoning was also significant to the .01 level, correlating .21. The remainder of the tests did not have significant validity coefficients.

**Detroit Mechanical Aptitude Test Total**: The test as a whole yielded a validity coefficient of .21 which is significant to the .01 level of confidence.

**Detroit Mechanical Aptitude Subtests**: Test #4, Arithmetic, is the only test which correlates significantly to the .01 level - being .22. Estimation of size, #3, and Mechanical Information, #6, correlate respectively .16 and .15 which are significant to the .05 level. The remainder do not correlate significantly.

**Summary**

(1) The Otis Thirty Minute version gave the highest correlation of any of the intelligence tests, being .27 and at the .01 level of confidence. The Wonderlic, however, was close behind with a correlation of .25. The other measures were significant at the .05 level.

(2) The Numbers test proved to be significant at the .01 level and also to be the best measure of prediction of the whole battery. The Reasoning Test, #3, has a
correlation figure consistent with the intelligence coefficients being at the .01 level with r of .21.

(3) The Detroit Mechanical Aptitude Test is significantly correlated with the criterion, and is at the .01 level of confidence with a coefficient of .21.

(4) The numerical test, #4 in the Detroit, like the Numbers Test in the PMA, yielded the highest correlation of the subtests - .22 - which is significant at the .01 level. Estimation of size and Mechanical Information were significant at the .05 level.

ii. Writers

Intelligence and Classification Tests: The Twenty Minute Otis yielded the highest correlation with the course marks-.35. Due to the size of the sample this is only significant to the .05 level, and the Otis Thirty Minute and the Wonderlic, which correlated .29 and .25 respectively, and the PMA and Navy "G" which correlated .27 and .19 respectively, are not significant.

Primary Mental Abilities Test: Only Test #4, Numbers, correlates relatively high, being .27 which is not significant. The other tests are very low.

Detroit Clerical Aptitude Test Total Score: This test gives a validity coefficient of .44 which is significant at the .05 level of confidence.
Detroit Clerical Aptitude Subtests: Subtest #2, Item checking, #3 - Arithmetic and #7 - Digit Alphabet Sorting, all yielded statistically significant coefficients of prediction, namely, .44, .32 and .56 respectively. The other correlations for the subtests were of a lower order and not significant.

Summary

(1) The Twenty Minute Otis yielded the highest coefficient of the intelligence and classification tests. There was again little difference between the Thirty Minute Otis and the Wonderlic and they were comparable in size to the coefficients obtained for these two tests with the Stokers. The total of the PMA was on a par with the Otis Thirty but once again the Navy "G" Test made the poorest showing.

(2) Again, as with the Stokers, the Numbers Test - #4 - proved to be the best measure of prediction of the PMA Tests.

(3) The Detroit Clerical Aptitude Test gave a prediction coefficient considerably higher than any of the intelligence tests.

(4) The Digit-Alphabet Sorting Test yielded the highest coefficient of any test or subtest for either the Stokers or Writers. Item Checking - #2 - and Arithmetic - #3 - were also significantly correlated with the course marks.
General Observations

The Writer group, being so small, does not allow for significant correlations, even when the coefficient is relatively high. Therefore it is necessary to consider correlations like .27 as having some potential significance even though it is not significant to the .05 level.

C. ANALYSIS OF TEST SCORES WITH OTIS S-A THIRTY MINUTE

The Writers and Stokers were combined into one group for the purposes of correlating all the other tests and subtests with the Otis S-A., with the exception of the correlation for the Aptitude Tests, as the Writers only took the Clerical Aptitude Test and the Stokers only took the Mechanical Aptitude Test. A tabulation of the test correlations from which the following analysis is made are to be found in Tables 7, 8 and 9.

Intelligence and Classification Tests: The Twenty Minute Otis correlates the highest with the Thirty Minute, the result being .89. The Navy "G", the Wonderlic and the PMA Total, in descending order, correlate .76, .73 and .61 respectively. These are all significant to the .01 level.

Primary Mental Abilities Tests: Reasoning - #3 - with a figure of .55 correlates the highest with the Otis, while Vocabulary - #1, and Word Fluency - #5, are somewhat lower, being of the order of .49 and .42. The Number test - #4 - is of a lower order, being only .29, but is also significant
TABLE VII

Correlation between scores obtained on the Otis Thirty Minute Test and the other tests and subtests in the battery for the combined sample of Stokers and Writers.

<table>
<thead>
<tr>
<th></th>
<th>Coeff. of Correlation</th>
<th>Signif. level</th>
<th>Number of Testees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otis Thirty Minute and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Otis Twenty Minute</td>
<td>.89</td>
<td>.01</td>
<td>119</td>
</tr>
<tr>
<td>Wonderlic</td>
<td>.73</td>
<td>.01</td>
<td>177</td>
</tr>
<tr>
<td>PMA Total Score</td>
<td>.61</td>
<td>.01</td>
<td>174</td>
</tr>
<tr>
<td>PMA #1, Verbal Meaning</td>
<td>.49</td>
<td>.01</td>
<td>174</td>
</tr>
<tr>
<td>PMA #2, Space</td>
<td>.16</td>
<td>.05</td>
<td>174</td>
</tr>
<tr>
<td>PMA #3, Reasoning</td>
<td>.55</td>
<td>.01</td>
<td>174</td>
</tr>
<tr>
<td>PMA #4, Number</td>
<td>.29</td>
<td>.01</td>
<td>174</td>
</tr>
<tr>
<td>PMA #5, Word Fluency</td>
<td>.42</td>
<td>.01</td>
<td>174</td>
</tr>
<tr>
<td>Navy G Classification</td>
<td>.76</td>
<td>.01</td>
<td>178</td>
</tr>
</tbody>
</table>
TABLE VIII

Correlation between the Otis S-A Thirty Minute Test and the Detroit Mechanical Aptitude Test and subtests -- Sample 145 Stokers.

<table>
<thead>
<tr>
<th>Otis Thirty Minute and</th>
<th>Coeff. of Correlation</th>
<th>Significance at level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detroit Mechanical Total</td>
<td>.57</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #1, Tool Inf.</td>
<td>.17</td>
<td>.05</td>
</tr>
<tr>
<td>Det. Test #2, X's in circle</td>
<td>.04</td>
<td>Not</td>
</tr>
<tr>
<td>Det. Test #3, Est. of size</td>
<td>.23</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #4, Arithmetic</td>
<td>.24</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #5, Space org.</td>
<td>.37</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #6, Mech. Info.</td>
<td>.45</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #7, Direct. of Pull.</td>
<td>.48</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #8, Digit-Alpha.</td>
<td>.22</td>
<td>.01</td>
</tr>
</tbody>
</table>
### TABLE IX

Correlation between the Otis S-A Thirty Minute Test and the Detroit Clerical Aptitude Test and subtests -- Sample 34 Writers.

<table>
<thead>
<tr>
<th>Otis Thirty Minute Test and</th>
<th>Coefficient of Correlation</th>
<th>Significance at level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detroit Clerical Total</td>
<td>.41</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #1, Handwriting</td>
<td>.12</td>
<td>Not</td>
</tr>
<tr>
<td>Det. Test #2, Item Checking</td>
<td>.06</td>
<td>Not</td>
</tr>
<tr>
<td>Det. Test #3, Arithmetic</td>
<td>.40</td>
<td>.01</td>
</tr>
<tr>
<td>Det. Test #4, X's in circle</td>
<td>.17</td>
<td>Not</td>
</tr>
<tr>
<td>Det. Test #5, Clerical info.</td>
<td>.36</td>
<td>.05</td>
</tr>
<tr>
<td>Det. Test #6, Space org.</td>
<td>.17</td>
<td>Not</td>
</tr>
<tr>
<td>Det. Test #7, Digit-Alpha.</td>
<td>.25</td>
<td>Not</td>
</tr>
<tr>
<td>Det. Test #8, Filing</td>
<td>.28</td>
<td>Not</td>
</tr>
</tbody>
</table>
to the .01 level of confidence. Space has the lowest correlation with .16.

Detroit Mechanical Total and Subtests: The Detroit Mechanical Total correlates .57 for Stokers which is significant to the .01 level. None of the subtests correlate as highly. Direction of Pulley - #7, and Mechanical Information - #6, are the highest, being .48 and .45 respectively. Space Organization - #5 - with .37; Arithmetic - #4 - with .24; Estimation of Size - #3 - with .23 and Digit-Alphabet Sorting - #8 - with .22, follow. These are all at the .01 level, and Tool Information - #1 - with .17 is at the .05 level. Putting X's in circles is not significantly correlated with the Otis.

Detroit Clerical Aptitude Total and Subtests: The total score on the test correlates .41 with the Otis which correlates at the .01 level. The only subtests to correlate at the .01 level are Arithmetic - #3 and Clerical Information - #5 which correlate at the .05 level with an r of .36. Digit-Alphabet Sorting - #7 - with .25, and Filing Test - #8 - are the only other tests which correlate above .20, but these are not at the .05 level of confidence.

Summary

(1) The Twenty Minute Otis correlates the highest with the traditional Otis with .89. The other intelligence and classification tests correlate sufficiently highly to indicate that they are measuring to a large extent the
same thing.

(2) As might be anticipated, the Reasoning Test - #3 - of the PMA correlates the highest with the intelligence test. The Number and Space Tests do not overlap to any extent with the Otis, correlating with it only .29 and .16 respectively.

(3) The Detroit Mechanical Tests Total Score shows considerable overlap in its correlation with the Otis, but nevertheless must be covering somewhat different areas. Only Mechanical Information - #6 - and Direction of Pulley - #7 - show any particular correspondence with the Otis.

(4) The Detroit Clerical Aptitude Test Total Score indicates some contiguity in measurement with the Otis, and of the subtests only Arithmetic and Clerical Information show similarity in scope of measurement. The remainder of the tests are relatively independent in measurement from the Otis.

D. ANALYSIS OF INTERCORRELATIONS OF TOTAL SCORES OF INTELLIGENCE AND CLASSIFICATION TESTS

All the intelligence tests and classification tests were intercorrelated, and the results of this are shown in Table 10. From this we can see the consistency with which the tests intercorrelate - that is to say, those tests which intercorrelate the highest with the Otis Thirty Minute tend to correlate
TABLE X

Correlation of total scores of intelligence and classification tests for 145 Stokers and 34 Writers.

<table>
<thead>
<tr>
<th></th>
<th>Otis 30 min.</th>
<th>Otis 20 min.</th>
<th>Wond.</th>
<th>Navy &quot;G&quot;</th>
<th>PMA Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otis 30 Min.</td>
<td>W --</td>
<td>.81</td>
<td>.64</td>
<td>.54</td>
<td>.59</td>
</tr>
<tr>
<td>Otis 20 Min.</td>
<td>R .95</td>
<td>--</td>
<td>.66</td>
<td>.45</td>
<td>.23</td>
</tr>
<tr>
<td>Wonderlic</td>
<td>I .68</td>
<td>.59</td>
<td>--</td>
<td>.61</td>
<td>.53</td>
</tr>
<tr>
<td>Navy &quot;G&quot;</td>
<td>T .71</td>
<td>.72</td>
<td>.58</td>
<td>--</td>
<td>.27</td>
</tr>
<tr>
<td>PMA Total</td>
<td>E .63</td>
<td>.49</td>
<td>.57</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

S
higher with the other tests than those which do not. It can be seen from either the W riters or Stokers matrix that a hierarchy exists and in the following descending order: Otis Thirty Minute, Otis Twenty Minute, Wonderlic, Navy "G", PMA Total.

E. ANALYSIS OF INTERCORRELATIONS OF TESTS OF THE PRIMARY MENTAL ABILITIES TEST

Intercorrelations between the various tests of the Primary Mental Abilities Test were worked out, and these were compared with a similar procedure done by Thurstone on 18,000 high school students. This data is presented in Table 11. As can be observed, only one correlation is significantly higher in this study than that of Thurstone's study, and most of the coefficients are lower. The average of Thurstone's intercorrelations was .214 as opposed to an average for this study of .158. As all Thurstone's coefficients are significant to the .01 level of confidence, then the lower correlations obtained in this study are therefore statistically lower, even though the groups are considerably smaller, rather than the lower coefficients being an artifact of the different size of the groups. Therefore, it can be stated that these tests are as independent measures or "factors" in this study as that claimed by Thurstone for his test as standardized.

F. ANALYSIS OF MULTIPLE CORRELATIONS

An analysis and evaluation of the tests could be made by
**TABLE XI**

Comparison of Intercorrelations obtained (a) by Thurstone (b) Present Study.

<table>
<thead>
<tr>
<th>Test</th>
<th>Verbal</th>
<th>Space</th>
<th>Reasoning</th>
<th>Number</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Thurstone</td>
<td>-</td>
<td>.176</td>
<td>.207</td>
<td>-</td>
<td>.248</td>
</tr>
<tr>
<td>(b) Hill</td>
<td>-</td>
<td>.055</td>
<td>.194</td>
<td>-</td>
<td>.178</td>
</tr>
<tr>
<td>Space</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Thurstone</td>
<td>-</td>
<td>.238</td>
<td>.207</td>
<td>-</td>
<td>.248</td>
</tr>
<tr>
<td>(b) Hill</td>
<td>-</td>
<td>.246</td>
<td>.194</td>
<td>-</td>
<td>.178</td>
</tr>
<tr>
<td>Reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Thurstone</td>
<td>-</td>
<td>.238</td>
<td>.207</td>
<td>-</td>
<td>.248</td>
</tr>
<tr>
<td>(b) Hill</td>
<td>-</td>
<td>.246</td>
<td>.194</td>
<td>-</td>
<td>.178</td>
</tr>
<tr>
<td>Number</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Thurstone</td>
<td>-</td>
<td>.238</td>
<td>.207</td>
<td>-</td>
<td>.248</td>
</tr>
<tr>
<td>(b) Hill</td>
<td>-</td>
<td>.246</td>
<td>.194</td>
<td>-</td>
<td>.178</td>
</tr>
</tbody>
</table>
calculating all the permutations and combinations of 17 tests with the criterion, in combinations of 2, 3, 4 and 5. This was of course beyond the scope of this project. Instead, an inspection method was utilized in which those tests and sub-tests were chosen which yielded higher correlations and yet, on the basis of correlation or rationale, were apparently more or less independent of each other. Further to this, a degree of uniformity was imposed which further limited the choice, in that the tests selected were the same for Stokers and Writers. Thus the multiple correlations with the criteria were the Otis S-A Thirty Minute, Primary Mental Abilities Test #4 (Number) and the total score for either the Detroit Clerical or the Mechanical Aptitude Test. For the Stokers a multiple correlation of .41 was obtained, and in the case of the Writers a multiple of .47 was obtained.

G. ANALYSIS OF SCATTERGRAMS

From the scattergrams it can be seen that the relationships on which the correlations depend is not curvilinear. From an analysis of the scattergram of the Writers it can readily be noted that the spread of the group is very narrow. In the case of the Stokers there is some disturbance in the correlations, occasioned by three testees who did relatively poorly on the course appraisal but did considerably better on most of the tests. These three then contributed considerably to the lowering of the correlations.
CHAPTER VII

DISCUSSION OF RESULTS

The correlations of validity were, on the whole, of a rather low order, .34 being the maximum for the Stokers and .56 for the Writers. In the military situation, where success on the course is the criterion, low order correlations are apparently usual. Both Vernon (148) and Guilford (62) state that the psychologist must be satisfied with low correlations, and that these have practical value for prediction. It should be pointed out that the correlations obtained under the conditions imposed by this study are probably even lower than what might be obtained under the military conditions to which Vernon and Guilford were referring. In this regard there are several factors at work which cause an attrition of the validity coefficients. For example, the homogeneity of the sample which was discussed in Chapter IV, B, and demonstrated in Chapter VI, A, no doubt lessened the correlations. It might be anticipated that with larger samples, and especially in wartime with
recruitment of more diverse personnel, that correlations under similar experimental conditions would be considerably larger. As for the Stoker group, it has been shown that three individuals consistently had the effect of working against the correlation. It is quite possible that one single factor might account for their low criterion score which was not operative for the tests, (e.g. AWOL, personality disturbance, sickness) and, if they were eliminated on such investigated grounds, the correlations for Stokers would be increased appreciably.

There emerges from the analysis a consistency to the correlation results which is desirable. This consistency is both an inner consistency and one with results of a similar nature achieved outside the specific situation here studied. Also, a rationale exists for the relative differences in size of the correlations. Evidence to support these statements is here-with supplied.

In the PMA the Numbers Test, and in the Detroit its counterpart Arithmetic, were the best predictive measures for both Writers and Stokers. This is not only consistency within the study but, as we shall demonstrate, is consistent with the findings of other investigators. It can be seen from Table 12 that Arithmetic was the best measure of prediction in the whole battery for Writers and Stokers in the Royal Navy, and Vernon (148) states that elementary arithmetic usually gives better correlations with success in mechanical trades and with seamen than either mechanical or intelligence tests. In the U.S. Navy, arithmetic gave the highest correlation with the criterion, as will be seen from Table 14.
TABLE XII

Mean Validity Coefficients of Selection Tests in Clerical, Mechanical, and Other Branches of the Royal Navy.\textsuperscript{x}

<table>
<thead>
<tr>
<th>Naval Branches</th>
<th>No. of Investig.</th>
<th>Matrix</th>
<th>Shipley</th>
<th>Bennett</th>
<th>Arith. Squares</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writers etc.</td>
<td>5</td>
<td>.37</td>
<td>.38</td>
<td>.19</td>
<td>.44</td>
<td>.18</td>
</tr>
<tr>
<td>Stokers etc.</td>
<td>18</td>
<td>.30</td>
<td>.27</td>
<td>.33</td>
<td>.38</td>
<td>.26</td>
</tr>
<tr>
<td>Leading Seamen etc.</td>
<td>11</td>
<td>.31</td>
<td>.36</td>
<td>.33</td>
<td>.37</td>
<td>.24</td>
</tr>
</tbody>
</table>

\textsuperscript{x} From: Personnel Selection in the British Forces, Vernon and Parry, p. 214.
### TABLE XIII

Validity Coefficients of the Army General Classification Test (U. S. Army) *

<table>
<thead>
<tr>
<th>Population</th>
<th>Criterion</th>
<th>N</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative Clerical Trainees</td>
<td>Grades</td>
<td>2947</td>
<td>.40</td>
</tr>
<tr>
<td>Clerical Trainees, AAF</td>
<td>Grades</td>
<td>123</td>
<td>.44</td>
</tr>
<tr>
<td>Clerical Trainees, Armored</td>
<td>Grades</td>
<td>119</td>
<td>.33</td>
</tr>
<tr>
<td>Airplane Mechanics</td>
<td>Grades</td>
<td>99</td>
<td>.32</td>
</tr>
<tr>
<td>Airplane Mechanics</td>
<td>Grades</td>
<td>3061</td>
<td>.35</td>
</tr>
<tr>
<td>Motor Mechanics</td>
<td>Grades</td>
<td>318</td>
<td>.69</td>
</tr>
<tr>
<td>Tank Mechanics</td>
<td>Grades</td>
<td>237</td>
<td>.33</td>
</tr>
<tr>
<td>Aircraft Armorers</td>
<td>Grades</td>
<td>1907</td>
<td>.40</td>
</tr>
<tr>
<td>Aircraft Welders</td>
<td>Grades</td>
<td>583</td>
<td>.26</td>
</tr>
<tr>
<td>Bombsight Maintenance</td>
<td>Grades</td>
<td>195</td>
<td>.31</td>
</tr>
<tr>
<td>Sheet Metal</td>
<td>Grades</td>
<td>764</td>
<td>.27</td>
</tr>
<tr>
<td>Teletype Maintenance</td>
<td>Grades</td>
<td>487</td>
<td>.20</td>
</tr>
<tr>
<td><strong>Average Correlation</strong></td>
<td></td>
<td></td>
<td>.36</td>
</tr>
</tbody>
</table>

* From: *New Methods in Applied Psychology*, p. 54
TABLE XIV

Comparison of mean validity coefficients for the tests of the Basic Battery, U.S. Navy, Forms 1 and 2 (correlations corrected for curtailment in ranges of talent).

<table>
<thead>
<tr>
<th>School</th>
<th>N</th>
<th>GCT</th>
<th>READ</th>
<th>ARITH</th>
<th>MAT</th>
<th>MK(M)</th>
<th>MK(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form 1</td>
<td>1480</td>
<td>.52</td>
<td>.52</td>
<td>.63</td>
<td>.52</td>
<td>.46</td>
<td>.39</td>
</tr>
<tr>
<td>Form 2</td>
<td>1176</td>
<td>.58</td>
<td>.42</td>
<td>.50</td>
<td>.52</td>
<td>.62</td>
<td>.50</td>
</tr>
<tr>
<td>Electrical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form 1</td>
<td>1747</td>
<td>.52</td>
<td>.52</td>
<td>.59</td>
<td>.44</td>
<td>.35</td>
<td>.49</td>
</tr>
<tr>
<td>Form 2</td>
<td>1062</td>
<td>.48</td>
<td>.37</td>
<td>.49</td>
<td>.48</td>
<td>.36</td>
<td>.62</td>
</tr>
<tr>
<td>Gunner's Mate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form 1</td>
<td>1677</td>
<td>.38</td>
<td>.39</td>
<td>.31</td>
<td>.28</td>
<td>.40</td>
<td>.43</td>
</tr>
<tr>
<td>Form 2</td>
<td>809</td>
<td>.38</td>
<td>.32</td>
<td>.31</td>
<td>.50</td>
<td>.56</td>
<td>.54</td>
</tr>
<tr>
<td>Torpedomen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form 1</td>
<td>880</td>
<td>.32</td>
<td>.35</td>
<td>.28</td>
<td>.27</td>
<td>.39</td>
<td>.35</td>
</tr>
<tr>
<td>Form 2</td>
<td>786</td>
<td>.29</td>
<td>.26</td>
<td>.24</td>
<td>.33</td>
<td>.54</td>
<td>.39</td>
</tr>
</tbody>
</table>

TABLE XV

Intercorrelations among the tests and subtests of the Basic Battery, U.S. Navy on a national sample of 500. x

<table>
<thead>
<tr>
<th></th>
<th>GCT</th>
<th>READ</th>
<th>ARI</th>
<th>MAT</th>
<th>MK(E)</th>
<th>MK(M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCT</td>
<td>-</td>
<td>.81</td>
<td>.69</td>
<td>.60</td>
<td>.53</td>
<td>.49</td>
</tr>
<tr>
<td>READ</td>
<td>-</td>
<td>.69</td>
<td>.56</td>
<td>.51</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>ARI</td>
<td>-</td>
<td>.61</td>
<td>.47</td>
<td>.53</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>MAT</td>
<td>-</td>
<td></td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK(E)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK(M)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

x From: Psychology and Military Proficiency, Charles W. Bray, p. 65.
It might also be pointed out that in the Royal Navy study a correlation of .38 was obtained for Stokers which is consistent with .34 for the Numbers Test in this study.

From an analysis of the Detroit Clerical Aptitude Tests one can see also certain consistencies. The highest correlation of validity obtained was that of the Digit-Alphabet Sorting which is a standard item common to many clerical aptitude tests. Item Checking, which is one of the oldest and most reliable of tests, also correlated higher with the criterion than most, and the case for Arithmetic has already been stated.

In the intercorrelations between the Otis S-A Thirty Minute and the PMA Tests, Reasoning yielded the highest correlation and, on the basis of rationale, one would expect it to correlate highly with intelligence measures. Vocabulary, which since the days of Binet has always been considered a stable and reliable predictor of intelligence, also correlates comparatively high in this study. On the other hand, tasks which one would not expect to have much relation to intelligence - such as putting X's in circles and handwriting - do not, in fact, correlate with intelligence as tested by the Otis.

Other evidences of consistency are also found. In Ghiselli's study (53) (Table 16) we find that for an analysis of 85 validity coefficients the mean validity was .35. The results obtained for the Otis, while somewhat lower, were of the same order. The U.S. Army AGCT Test yielded somewhat superior coefficients also for clerical trainees in various branches, ranging from .33 to .44, as opposed to .29 validity coefficient for the Otis in
### TABLE XVI

Validity coefficients of unspecified intelligence tests for various types of employment.  

<table>
<thead>
<tr>
<th>Occupational Group</th>
<th>No. of Validity Coefficient</th>
<th>Mean Validity Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clerical workers</td>
<td>85</td>
<td>.35</td>
</tr>
<tr>
<td>Supervisors</td>
<td>9</td>
<td>.40</td>
</tr>
<tr>
<td>Salesmen</td>
<td>4</td>
<td>.30</td>
</tr>
<tr>
<td>Sales clerks</td>
<td>18</td>
<td>.09</td>
</tr>
<tr>
<td>Protective service</td>
<td>6</td>
<td>.25</td>
</tr>
<tr>
<td>Skilled workers</td>
<td>45</td>
<td>.55</td>
</tr>
<tr>
<td>Semi-skilled workers</td>
<td>13</td>
<td>.20</td>
</tr>
<tr>
<td>Unskilled workers</td>
<td>12</td>
<td>.08</td>
</tr>
</tbody>
</table>

*From: Ghisselli*
this study (Table 13). From Table 14 we see that a correlation of .60 was obtained between the U.S. Navy GCT and Mechanical Aptitude Test, whereas in this study a coefficient of .57 was obtained in a correlation between the Otis and the Detroit Mechanical Aptitude Test.

The only difficulties found in test administration were found with the Primary Mental Abilities Test and, in particular, with the Space Test. There seemed to be some confusion following the standardized presentation and instructions in each administration. This undoubtedly explains the loading in the first step interval as indicated in the psychogram Appendix D, fig.7, for Stokers and Writers. As for the administration as a whole, it might be possible that the low correlations are a result of the administration, but it can be seen that the Navy "G" Test, which was not administered by the group, was even lower than the other intelligence tests.

While the cut-off point for the Navy "G" Test was 36 it can be seen that some of the testees have scores below this. The explanation here is that they were retested and on the retest obtained the critical score or better. For the sake of uniformity the first score obtained was used for all testees who were retested. The Navy "G", having already been applied to the samples at enlistment and a selection made at that time, affected the tests' later usefulness particularly for prediction of success on courses. From Appendix we can see that the distribution of scores was unsatisfactory for this use, and
from a study of the validity coefficients obtained (Tables 5 and 6) this is borne out.

The Primary Mental Abilities Test was included in the battery to discover whether it might yield higher coefficients of validity and also whether it might have some diagnostic value. In this study it was found to operate somewhat differentially for Writers and Stokers. In the first place the Number Test yielded the highest correlation of any of the PMA Tests for both groups. In the case of the Stokers it yielded the highest correlation of any test, and therefore can be said to have some predictability. The Reasoning Test also has some prediction value for Stokers and, presumably, if the PMA was used as a selection device the performance on these tests would be most important. With the Writers only the Number Test yielded any correlation and this was .27 which is not significant. Therefore the test was not demonstrated to be as useful a predictive instrument for Writers.

The experiment was designed to investigate, among other things, the relative value of aptitude tests as opposed to intelligence tests. The Detroit Aptitude Tests were chosen as they contain a wider range of separate aptitude tests. It can readily be seen from the foregoing analysis that many of these tests have no predictive value whatsoever in the study (Table 6). However, the test as a whole was demonstrated as a better prediction measure for success on the Writers' course than any of the intelligence tests. For the Stokers, the Mechanical Aptit-
ude Test was not as comparably superior to the intelligence tests. The validity correlation was somewhat lower than that for the intelligence tests, particularly the Otis (.27 to .21). The test correlates .57 with the Otis and therefore has some independence which should increase the predictive value of a multiple containing both measures.
(1) The correlations of validity were of a relatively low order but nevertheless these coefficients, while not desirable, still have practical value and meaning. Therefore these tests will continue to be used until something superior is designed. An additional validity of the tests is suggested in the consistency of the scores, and it is felt that comparisons of these measures are legitimate.

(2) The low validity correlations are probably a result of the criterion and sample. The criterion, as was pointed out, was the best available measure. It yielded a reasonable distribution but was too restricted in dispersion and this was probably the result of a sample which had many restrictions placed upon it.
(3) The validity correlations for the Wonderlic were substantially almost as high as those for the other measures of intelligence in the study and, in consideration of the abbreviated testing time, it was assessed as the most useful (in the military situation) of the measures of intelligence studied.

(4) There can be no doubt that the Clerical Aptitude Test was a better measure than the intelligence scores for predicting success on the Writers' course. With the Stokers the case is not as clear cut, but there are indications that, used in combination with an intelligence test, it makes a contribution to increased proficiency in prediction. It was also indicated that many of the tests are useless for prediction in this situation. In particular this would include putting X's in circles, handwriting grades and filing.

(5) In the Primary Mental Abilities Test, the Number Test was the best measure of prediction for any test or subtest administered to the Stokers. Reasoning also had some predictive value. As for the Writers, only the Number Test had any value for prognosticating standing on the course. The Space Test suffered from some fault in the administration instructions which led in all cases of administration to confusion. Consequently the test itself has not really been investigated and might have value if administered with more comprehensible instructions.
The Navy "G" Test did not show up to any advantage and this is, in part at least, due to its previous screening use at recruitment of the sample. Its value for screening was not investigated by this study, but at least any extension of its use to other tasks can certainly be questioned on the basis of its performance in this study.

By calculating a multiple correlation, the low validity correlations can be raised for both Stokers and Writers to a more acceptable validity coefficient. The resultant coefficients that were obtained were .41 for Stokers and .47 for Writers. If these were again raised by manipulation for restriction of range, then it could be seen that a battery of an intelligence test, an arithmetic test and a specific aptitude test could yield a reasonably good measurement of prediction. The statistical manipulation for restriction for range was not done because it was felt that many of the restrictions found would always be operative in the military situation, and that the present coefficients have more practical meaning.

Certain tests were found to yield unsatisfactory score distributions, being in particular too difficult for the group (Stokers - Detroit Aptitude #5, PMA #2; Writers - PMA #2). However other tests were approaching either the upper or lower ends of the range and with larger samples might not be found satisfactory for certain uses.
CHAPTER IX

IMPLICATIONS ARISING OUT OF THE STUDY

(1) There is evidence to support the belief that testing programs might be shortened in time without any undue loss in validity. This particularly applies to intelligence tests but also to the "dead" subtests in the aptitude tests. Certainly this question merits further investigation.

(2) The value of the Navy Test, other than for a recruiting "screening" device, is questionable, and it should not be used otherwise without investigation of its merit for other jobs.

(3) It is recommended that a study of various criteria in the services be made. In this particular instance it may well be that success on the course is not particularly related to success on the job. Until a thorough
study is made all research in this field will be vulnerable to criticism in terms of criteria.

(4) While the Primary Mental Abilities Test did not demonstrate any marked superiority in this study, this is not necessarily a reflection on the methodological approach to psychometrics. It is quite possible that the "factors" involved in the test are not paramount in passing the examinations. From an examination of the literature it is felt that the approach used in the Guilford-Zimmerman Aptitude Survey (61) is the most promising. In brief, they recommend that certain fields of aptitudes, such as clerical and mechanical, be factor-analyzed and the primary factors isolated and tests constructed involving these. They have already prepared some tests along these lines, but many "factors" remain untouched. At the same time the specific job, for example - Stoker, should also be analyzed in terms of these "factors" to ascertain their respective loading. Then, when personnel are tested they are chosen on the basis of having psychograms matching those of the job.

(5) A long term study, or follow-up on this particular study, may quite possible demonstrate that many of the devices used here have a totally different value over time, and this should be borne in mind in accepting the foregoing conclusions.
BIBLIOGRAPHY


27. BRIGGS, H.L. What attainable standards are needed to measure results in vocational education and allied training? Voc. Guid. Mag., 1929, 7, 254-257.


47. FITTS, P.M. German applied psychology during World War I. Amer. Psychol., 1946, 5, 151-161.
64. GUILFORD, J.P. Some lessons from aviation psychology. Amer. Psychol., 1948, 3, 3-11.
71. HAY, E.N. The use of psychological tests in selection and promotion. Personnel, 1940, 16, 114-123.


79. HUNTER, W.S. Psychology in the War. Amer. Psychol., 1946, 1, 479-492.


100. OHMANN, O.A. The measurement of capacity for skill in stenography. Psychol. Monogr., 1926, 36, 54-70.


103. OSRD Validity of an experimental battery of aptitude tests at the Ordnance and Gunnery School. 1944, Publ. Board No. 13299, pp. 36.


113. ROBINSON, B.W. An experimental study of certain tests as measures of natural capacity and aptitude for typewriting. _Psychol. Monogr._, 1926, 36, 38-53.

114. RODGER, A. The work of the Admiralty psychologists. _Occup. Psychol._, 1945, 19, 132-139.


121. SLATER, P. Some group tests of spatial judgment or practical ability. *Occup. Psychol.*, Lond., 1940, **14**, 39-55.


133. THURSTONE, L.L. Psychological implications of factor analysis. Amer. Psychol. 1948, 3, 402-408.


139. TRAXLER, A.E. Correlations between mechanical aptitude scores and mechanical comprehension scores. Occupations, 1943, 22, 42-43.


141. TUCKMAN, J. The correlations between mechanical aptitude and mechanical comprehension scores. Occupations, 1944, 22, 244-245.


152. WATSON, K.B. Prognostic value of psychological tests in the Navy Training Program. *Amer. Psychol.*, 1946, 1, 446.


162. YUM, K.S. Primary Mental abilities and scholastic success in the divisional studies at the University of Chicago. *Psychol. Bull.*, 1941, 38, 552.
### APPENDIX A

### ASSESSMENT SCALE

<table>
<thead>
<tr>
<th>NAME</th>
<th>RATING</th>
<th>O.N.</th>
</tr>
</thead>
</table>

1. **CLEANLINESS AND APPEARANCE OF PERSON AND BARRACKS**
   - Is the man always clean and neatly dressed? Is his hair trimmed and combed? Is he well shaved? Does he have a military bearing?
   - Are his locker and gear clean and in perfect order? Does he contribute to the cleanliness of the barracks?

<table>
<thead>
<tr>
<th>EXCELLENT</th>
<th>GOOD</th>
<th>AVERAGE</th>
<th>FAIR</th>
<th>POOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>4.0</td>
<td>3.5</td>
<td>2.5</td>
<td>2.0</td>
</tr>
</tbody>
</table>

2. **MILITARY COURTESY**
   - Does the man salute smartly and carefully at all required times?
   - Does he say "Sir" habitually when talking to an officer? Is he cheerful in taking orders and carrying out his duties?
   - Does he conduct himself as a gentleman at all times?

<table>
<thead>
<tr>
<th>EXCELLENT</th>
<th>GOOD</th>
<th>AVERAGE</th>
<th>FAIR</th>
<th>POOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>4.0</td>
<td>3.5</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

3. **COOPERATION AND TEAMWORK**
   - Does the man always participate in group activities?
   - Does he help others without being ordered to do so?
   - Does he do more than his share when working in a group?
   - Does he know how to work with others?
   - Does he subordinate his own wishes to the welfare of the group?
   - Is he loyal to his shipmates and his superiors?

<table>
<thead>
<tr>
<th>EXCELLENT</th>
<th>GOOD</th>
<th>AVERAGE</th>
<th>FAIR</th>
<th>POOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>4.0</td>
<td>3.5</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

4. **STRAIGHTFORWARDNESS**
   - Is the man always truthful?
   - Does he respect the rights and property of others?
   - Can he be counted on to do the honest thing?

<table>
<thead>
<tr>
<th>EXCELLENT</th>
<th>GOOD</th>
<th>AVERAGE</th>
<th>FAIR</th>
<th>POOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>4.0</td>
<td>3.5</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

5. **RESPONSIBILITY**
   - Does the man always carry out orders with precision and despatch?
   - Is he answerable for all assignments?
   - Does he perform every task without being checked on?
   - Can he be relied upon?

<table>
<thead>
<tr>
<th>EXCELLENT</th>
<th>GOOD</th>
<th>AVERAGE</th>
<th>FAIR</th>
<th>POOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>4.0</td>
<td>3.5</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

6. **INITIATIVE**
   - Does the recruit observe things to be done and do them without being told to do so?
   - Is he a self-starter?

<table>
<thead>
<tr>
<th>EXCELLENT</th>
<th>GOOD</th>
<th>AVERAGE</th>
<th>FAIR</th>
<th>POOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>4.0</td>
<td>3.5</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

7. **KNOWLEDGE AND PERFORMANCE**
   - Does the man give evidence of possessing complete information on the subjects covered?
   - Has he acquired the skills taught?
   - Can he do most of the required tasks?

<table>
<thead>
<tr>
<th>EXCELLENT</th>
<th>GOOD</th>
<th>AVERAGE</th>
<th>FAIR</th>
<th>POOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>4.0</td>
<td>3.5</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>
APPENDIX B
Stokers' Examination Form

NAME

RATE

O.N.

CLASS NO.

(15) 1. Sketch a single type Evaporater and Distiller Plant.

(10) 2. Where would you find the following pieces of machinery? What are their uses?

1. Air Ejector.
2. Extraction Pump.
3. Evaporater.
5. Condensor.
6. Distiller.
7. Auxiliary Feed Pump.
8. Feed Water Heater.

(10) 3. List any ten Mountings on a Yarrow Boiler.

(10) 4. How would you test a boiler gauge glass to make sure the true water level is showing?

(10) 5. True or False:-

1. Transfer pump pumps oil through the lub. oil system.
2. Extraction pump pumps water directly to the boiler.
3. Safety lamp is for testing confined spaces.
4. Feed Pump pumps water to the fire main.
5. Main circulator - to supply the Evaporater.
7. Auxiliary Exhaust is used in the feed regulator.
8. Drain Cooler - used to cool the auxiliary exhaust.
9. Plummer Blocks - carry the weight of the prop- ellor shaft.
10. Evaporater Plant - to make fresh water.

(10) 6. How many boilers in a:-

1. Cruiser
2. Destroyer (Tribal Class)
3. Air Craft Carrier

(10) 7. What are the causes of :

1. Black Smoke
2. White Smoke
3. Pulsation
4. Back Flash
5. Sprayers becoming extinguished.

(10) 8. What would happen if :

1. Boiler short of water
2. Boiler with too much water

(10) 9. Clean a Yarrow Boiler Internally

10. Five Marks for Books.
APPENDIX C

Intercorrelations of certain tests and subtests, classification tests and criterion for sample of 154 Stokers and 39 Writers

<table>
<thead>
<tr>
<th></th>
<th>CRIT. CTIS 30</th>
<th>CTIS 20</th>
<th>T 1</th>
<th>T 2</th>
<th>T 3</th>
<th>T 4</th>
<th>T 5</th>
<th>T 6</th>
<th>T 7</th>
<th>T 8</th>
<th>&quot;H&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRITERION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.27</td>
<td>.16</td>
<td>.25</td>
<td>.16</td>
<td>.10</td>
<td>.07</td>
<td>.21</td>
<td>.34</td>
<td>.02</td>
<td>.21</td>
<td>.12</td>
</tr>
<tr>
<td>Writers</td>
<td>.29</td>
<td>.35</td>
<td>.25</td>
<td>.27</td>
<td>.10</td>
<td>.13</td>
<td>.10</td>
<td>.27</td>
<td>-.02</td>
<td>.44</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTIS 30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.81</td>
<td>.64</td>
<td>.59</td>
<td>.22</td>
<td>.41</td>
<td>.12</td>
<td>.04</td>
<td>.23</td>
<td>.24</td>
<td>.37</td>
<td>.45</td>
</tr>
<tr>
<td>Writers</td>
<td>.95</td>
<td>.68</td>
<td>.63</td>
<td>.14</td>
<td>.12</td>
<td>.06</td>
<td>.40</td>
<td>.17</td>
<td>.36</td>
<td>.17</td>
<td>.25</td>
</tr>
<tr>
<td>Total</td>
<td>.89</td>
<td>.73</td>
<td>.61</td>
<td>.49</td>
<td>.16</td>
<td>.55</td>
<td>.29</td>
<td>.42</td>
<td></td>
<td></td>
<td>.76</td>
</tr>
<tr>
<td>OTIS 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.66</td>
<td>.59</td>
<td>.53</td>
<td>.29</td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.69</td>
<td>.49</td>
<td>.49</td>
<td>.39</td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WONDERLIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.53</td>
<td>.53</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.57</td>
<td>.57</td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMA TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.52</td>
<td>.52</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.49</td>
<td>.49</td>
<td>.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMA 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.06</td>
<td>.06</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.09</td>
<td>.09</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMA 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.13</td>
<td>.13</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.21</td>
<td>.21</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMA 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMA 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.19</td>
<td>.19</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMA 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.13</td>
<td>.13</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writers</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stokers</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.19</td>
<td>.19</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DET. APT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.41</td>
<td>.41</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAVY &quot;G&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX  D

HISTOGRAMS

of

SCORE DISTRIBUTIONS
Fig. 1  Distribution of Scores of 144 Stokers
On Criterion Course Marks

Fig. 2  Distribution of Scores of 145 Stokers
On 30 Minute Otis S-A.
Fig. 3 Distribution of Scores of 90 Stokers
On Otis S-A 20 Minute

Fig. 4 Distribution of Scores of 143 Stokers
On Wanderlic Test
Fig. 5  Distribution of Scores of 140 Stokers
          On PMA Total Scale

Fig. 6  Distribution of Scores of 140 Stokers
          On PHA Test #1
Fig. 7 Distribution of Scores of 140 Stokers On PMA Test #2

Fig. 8 Distribution of Scores of 140 Stokers On PMA Test #3
Fig. 9  Distribution of Scores of 138 Stokers on P.M.A. Test #4

Fig. 10  Distribution of Scores of 139 Stokers on P.M.A. Test #5
Fig. 11  Distribution of Scores of 145 Stokers

Fig. 12  Distribution of Scores of 145 Stokers
On Detroit Apt. Test No. 1
Fig. 13 Distribution of Scores of 145 Stokers On Detroit Mech Apt. Test #2

Fig. 14 Distribution of Scores of 145 Stokers On Detroit Mech Apt. Test #3
Fig. 15  Distribution of Scores of 146 Stokers
On Detroit Mech. Apt. Test # 4

Fig. 16  Distribution of Scores of 146 Stokers
Fig. 17 Distribution of Scores of 145 Stokers

Fig. 18 Distribution of Scores of 145 Stokers
Fig. 19  Distribution of Scores of 145 Stokers

Fig. 20  Distribution of Scores of 145 Stokers
On N A VY ' G ' Classification Test
Fig. 21 Distribution of Scores of 34 Writers On Criterion Course Marks

Fig. 22 Distribution of Scores of 34 Writers On Otis S-A 30 Minute Test
Fig. 23 Distribution of Scores of 29 Writers
On Otis S-N 20 Minute Test.

Fig. 24 Distribution of Scores of 34 Writers
On Wonderlic Test.
Fig. 25 Distribution of Scores of 34 Writers
On PMA Total Score.

Fig. 26 Distribution of Scores of 34 Writers
On PMA Test #1
Fig. 17  Distribution of Scores of 34 Writers
On PMA Test #2

Fig. 18  Distribution of Scores of 34 Writers
On PMA Test #3
Fig. 19  Distribution of Scores of 34 Writers
On PMA Test #4

Fig. 30  Distribution of Scores of 34 Writers
On PMA Test #5
Fig. 31 Distribution of Scores of 34 Writers
On Detroit Clerical Apt. Total Score

Fig. 32 Distribution of Scores of 34 Writers
On Detroit Clerical Apt. Test #1
Fig. 33 Distribution of Scores of 34 Writers
On Detroit Apt. Test, Clerical. ^#2

Fig. 34 Distribution of Scores of 34 Writers
On Detroit Clerical Apt. Test #3
Fig. 35: Distribution of Scores of 34 Writers on Detroit Clerical Apt. Test #4

Fig. 36: Distribution of Scores of 34 Writers on Detroit Clerical Apt. Test #5
Distribution of Scores of 34 Writers on Detroit Clerical Apt. Test 

Fig. 37

Distribution of Scores of 34 Writers on Detroit Clerical Apt. Test 

Fig. 38
Fig. 39
Distribution of Scores of 34 Writers
On Detroit Clerical Apt. Test *8

Fig. 40
Distribution of Scores of 33 Writers
On Navy "G" Classification Test