

Armed Services Technical Information Agency

Because of our limited supply, you are requested to return this copy WHEN IT HAS SERVED YOUR PURPOSE so that it may be made available to other requesters. Your cooperation will be appreciated.

AD

46265

NOTICE: WHEN GOVERNMENT OR OTHER DRAWINGS, SPECIFICATIONS OR OTHER DATA ARE USED FOR ANY PURPOSE OTHER THAN IN CONNECTION WITH A DEFINITELY RELATED GOVERNMENT PROCUREMENT OPERATION, THE U. S. GOVERNMENT THEREBY INCURS NO RESPONSIBILITY, NOR ANY OBLIGATION WHATSOEVER; AND THE FACT THAT THE GOVERNMENT MAY HAVE FORMULATED, FURNISHED, OR IN ANY WAY SUPPLIED THE SAID DRAWINGS, SPECIFICATIONS, OR OTHER DATA IS NOT TO BE REGARDED BY IMPLICATION OR OTHERWISE AS IN ANY MANNER LICENSING THE HOLDER OR ANY OTHER PERSON OR CORPORATION, OR CONVEYING ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE OR SELL ANY PATENTED INVENTION THAT MAY IN ANY WAY BE RELATED THERETO.

Reproduced by
DOCUMENT SERVICE CENTER
KNOTT BUILDING, DAYTON, 2, OHIO

UNCLASSIFIED

AD No 46265

ASTIA FILE COPY

**BUREAU OF NAVAL PERSONNEL
TECHNICAL BULLETIN 54-11**

**SELECTION AND CLASSIFICATION
TESTS FOR WOMEN**

A Review of the Literature

Barbara Wand
and
William G. Mollenkopf

Educational Testing Service
under
Contract No. Nonr-694(00)

**Classification & Survey Research Branch
Personnel Analysis Division**

1 June 1954



Bureau of Naval Personnel

TECHNICAL BULLETIN 54-11

SELECTION AND CLASSIFICATION TESTS FOR WOMEN

A Review of the Literature

Barbara Wand

and

William G. Mollenkopf

Educational Testing Service

under

Contract No. Nonr-694(00)

Classification & Survey Research Branch

Personnel Analysis Division

1 June 1954

SELECTION AND CLASSIFICATION TESTS FOR WOMEN:

A REVIEW OF THE LITERATURE

Abstract

The purpose of this review of the literature was to examine the appropriateness of using with women the same selection and classification procedures that are used with men in situations in which both men and women are selected for the same jobs. Particular attention was paid to reports of the selection of women for jobs similar to billets in the U. S. Navy.

Two valuable sources of information on women which included comparable data for men were U. S. Air Force reports of the selection of personnel for Air Force technical schools and British reports of the selection of women for the Auxiliary Territorial Service during World War II.

Most industrial studies were based on samples of one sex. It became apparent that in normal times most civilian jobs are held predominantly by members of one sex or the other, and hence the practical problem of the influence of sex differences on predictive measures has not been investigated to any extent by those who are conducting personnel research in industry.

The findings tend to support the assumption that tests developed and used for the selection of men must be carefully examined prior to their use in selecting women for the same jobs. This seems particularly true for tests in mechanical and computational areas. In general, findings seem to indicate that a given tests score may not predict the same level of on-the-job performance for a woman as for a man, there being evidence that women do better than would be predicted from scores based on tests and procedures developed for men.

SELECTION AND CLASSIFICATION TESTS FOR WOMEN:

A REVIEW OF THE LITERATURE

Table of Contents

	<u>Page</u>
Abstract	1
I. Introduction	1
A. General Setting	3
B. Organization	7
II. Navy Tests Administered to Enlisted Women	10
III. Selection and Classification of Women in the British Armed Forces	13
IV. Performance of Women in Air Force Technical Schools	25
V. Civilian Studies of the Selection of Women for Navy-like Jobs Using Tests Like Those in the Navy Basic Test Battery	32
A. Clerical Workers	32
B. Industrial Workers	39
VI. Civilian Studies of the Selection of Women for Navy-like Jobs Using Tests Unlike Those in the Navy Basic Test Battery	41
A. Clerical Workers	41
B. Industrial Workers	41
VII. Sex Differences in Scores on Tests Like Those in the Navy Basic Test Battery	44
A. Clerical Tests	44
B. Tests of Mechanical Comprehension	46
C. Tests of Intelligence	47
VIII. Implications	48
IX. Bibliography	52

SELECTION AND CLASSIFICATION TESTS FOR WOMEN:

A REVIEW OF THE LITERATURE

I. Introduction

During any period of national mobilization it seems likely that women will be called upon to carry out a large number of duties and to fill a large number of jobs ordinarily assigned to men in peacetime. This greater use of women might be regarded by many persons as an emergency measure; forced upon us by necessity; these persons might argue that women constitute a labor group whose proficiency on many jobs is generally inferior to that of men. However, it is difficult to reconcile this viewpoint with the increasing peacetime employment of women in non-clerical positions.

As can be seen by examining Table 1, women have become an increasingly larger part of the labor force of the United States over the past seventy years. Whereas in 1880 only 15 per cent of persons gainfully employed were women, by 1953 women made up 30 per cent of this group. Census data indicate that in the last quarter of a century the distribution of women among the various occupational groups has changed steadily. In mechanical and manufacturing industries, for instance, the proportion of the workers who were women increased from 13 per cent in 1930 to 19 per cent in 1953. Furthermore, a greater proportion of mechanical and manufacturing jobs are today filled by women: The percentage of the total number employed on these jobs who were women increased from 18 per cent in 1930 to 23 per cent in 1953.

The United States Navy during World War II found personnel of its female component, the WAVES, to be effective and competent in a considerable number of billets ranging from draftsman to disbursing clerk. In the selection and classification of WAVES personnel, both officer and enlisted, the tests that were and are being employed have largely been tests developed for use with male personnel. That these tests might not be entirely appropriate for use with women has been a matter of concern to the Navy. In particular, it seemed necessary to examine the implication of the fact that women obtain lower scores on some aptitude tests than men do, notably on tests in the mechanical or quantitative area.

Table 1

Proportion of Women in the Total Group
of Gainfully Employed Persons in the United States
(Over 10 years of age between 1880 and 1930
and over 14 years of age thereafter)*

<u>Year</u>	<u>Per Cent</u>
1880	14.7
1890	17.4
1900	18.8
1910	23.4
1920	21.1
1930	22.0
1940	24.9
1950	28.9
1953	30.1

*Compiled from the World Almanac and Book of Facts:
New York: New York World Telegram and Sun. 1935,
p. 314; 1945, p. 546; 1951, p. 582; 1954, p. 261.
Primary Source: Bureau of the Census.

(In order not to leave the impression that women generally make lower scores on tests than men, it should be mentioned that women do as well, if not better, than men on verbal and clerical tests.)

In the spring of 1953 the Bureau of Naval Personnel, through contract between the Office of Naval Research and the Educational Testing Service, arranged to have a project carried out that would deal with the appropriateness of existing selection and classification tests for use with enlisted women in the Navy. As one phase of this project there was to be a critical review of the psychological literature on the measurement and evaluation of aptitudes and skills of women both in military and in industrial settings comparable to those to which enlisted women in

the Navy might be assigned.

A. General Setting

The most relevant evidence for evaluating the appropriateness of selection and classification tests for use with women is provided by the relationship between scores on the test and a suitable assessment of performance on the job. (Since the latter constitutes the pay-off, we shall refer to it as the "criterion.") Inasmuch as most of the existing Navy tests were developed for men and have been used with women in the same way as with men, the criterion-test relationship for women should be compared with that for men in order to arrive at conclusions regarding the appropriateness of the test procedures for use with women.

As has been mentioned, on some tests enlisted women in the Navy have on the average scored significantly lower than have enlisted men. This finding taken alone is ambiguous in meaning. The significance of a given score may or may not be the same for a woman as for a man.

A few diagrams may serve to illustrate this point. In the first three of these, various possible relationships between the criterion and the test score have been depicted. In each case the mean score observed for women is taken to be lower than that for men. In Figure 1, the regression of criterion performance on test score has been illustrated as being the same for women as for men. This illustration would correspond to (and justify) the practice of treating a given score as having identical meaning for a woman as for a man. According to this representation women score lower than men on the test -- and they do not, on the average, perform as well on the job. If actual evidence from studies could be found to fit this model, such evidence would support use of the same tests with both sexes, with the same cutting score being appropriate for both.

But next consider Figure 2. Here, women are again shown as generally scoring lower on the test than men, but this time they are represented as, on the average, performing as well on the job as do the men. A given score does not, in this model, have the same implication for a woman that it has for a man. While the test is effective for

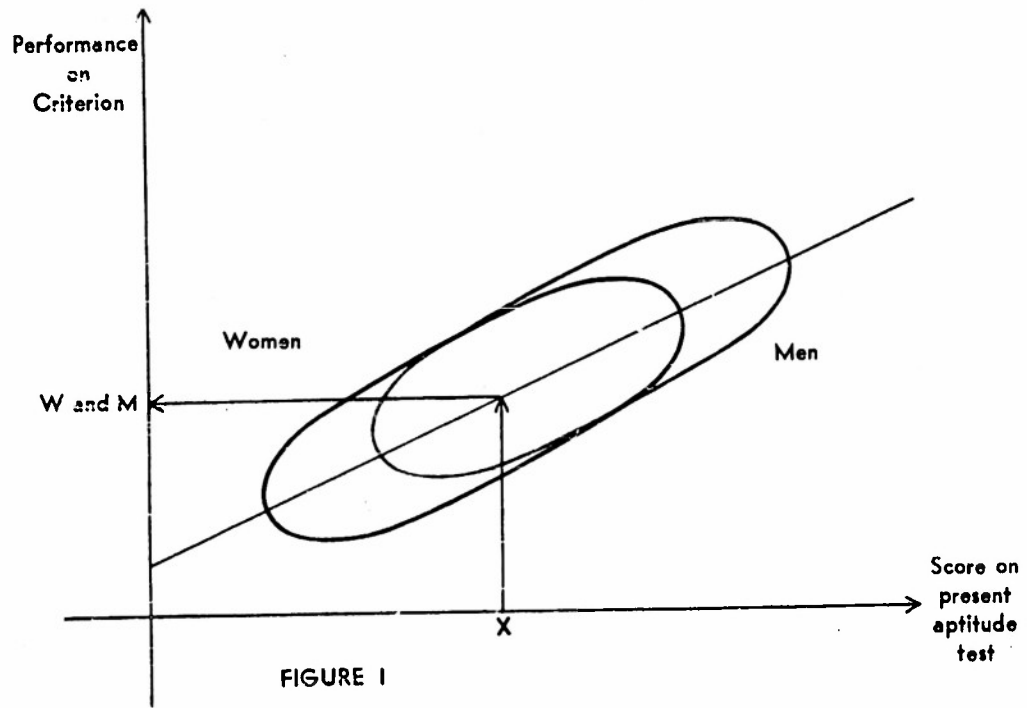


FIGURE 1

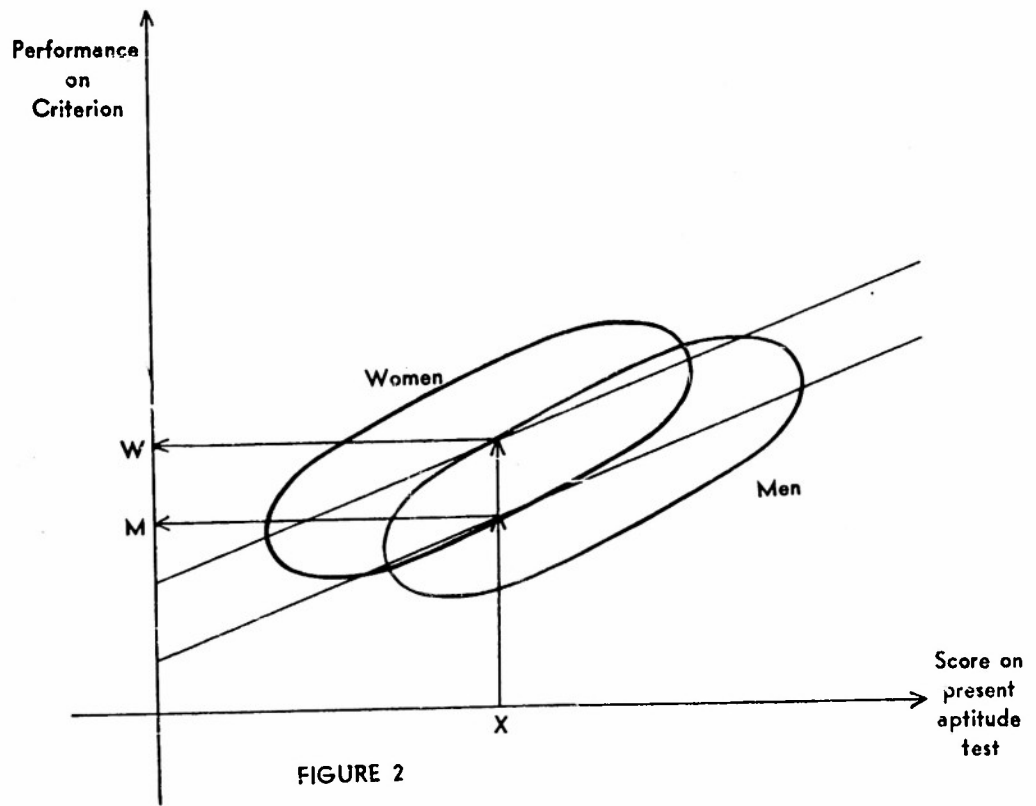
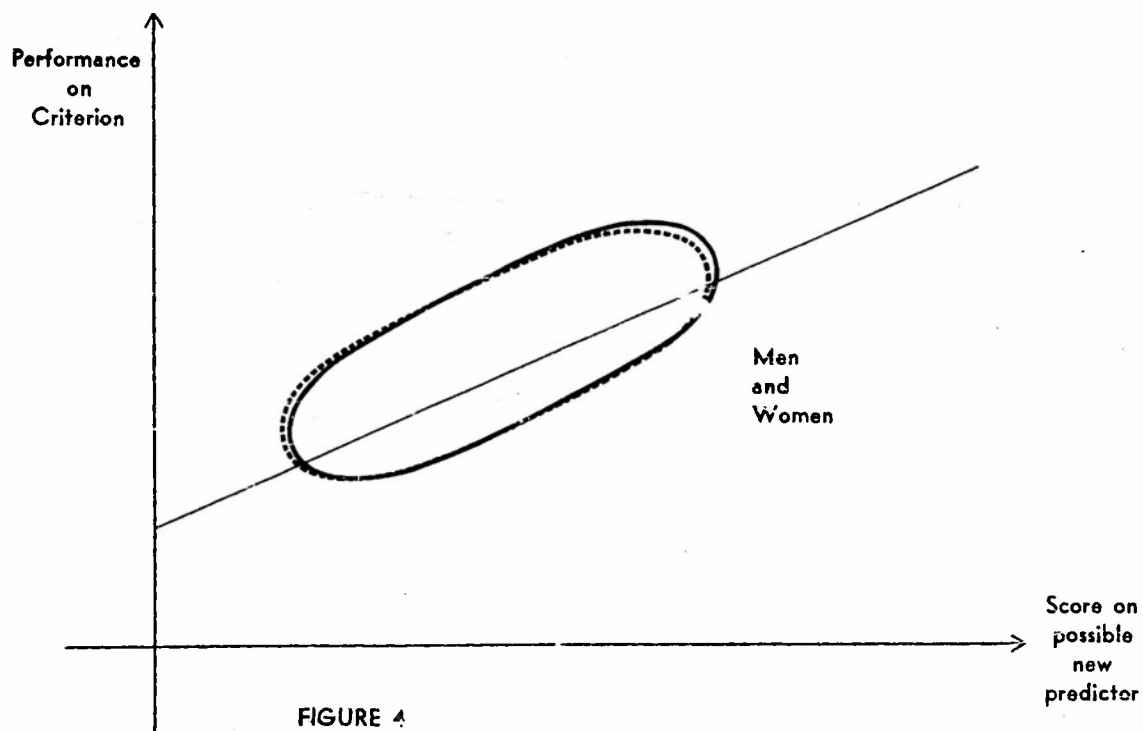
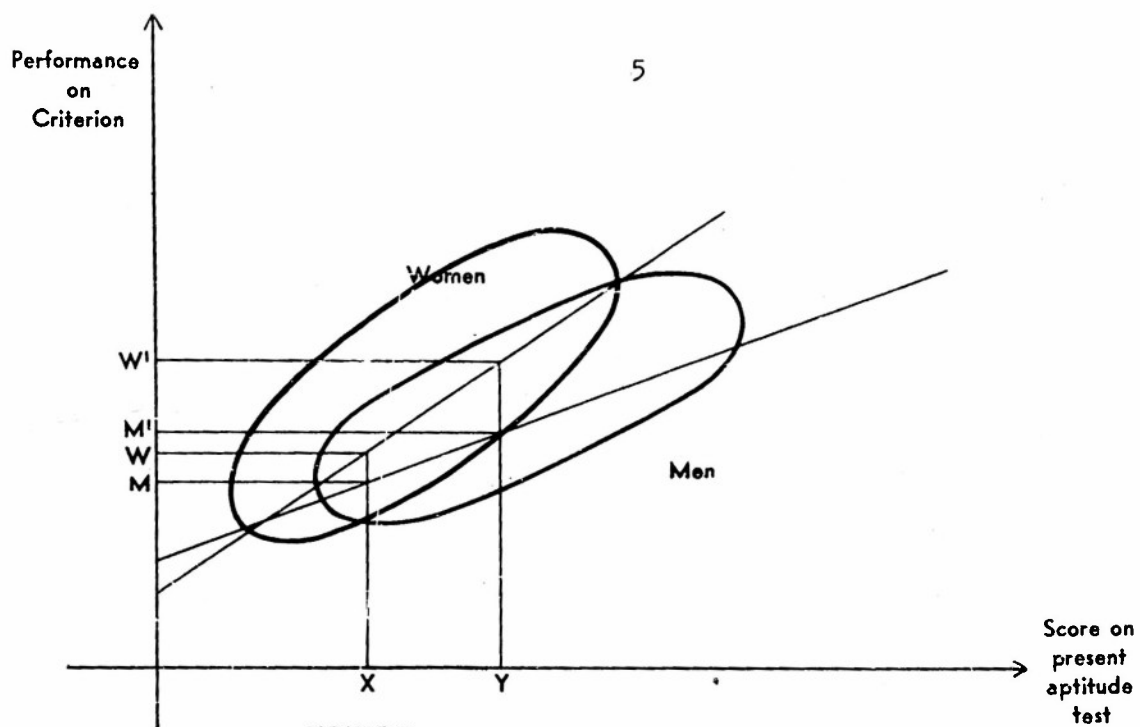


FIGURE 2



identifying the better and poorer prospects in each group, for adequate interpretation of a given score the sex of the recruit must be taken into account. A recruit with a score of X points on the test will be most likely to have a criterion performance of W if a female and M if a male. Operationally, if evidence supported this model, different procedures would be required for women than for men. One cutting score for women and another for men would probably be the simplest way of taking the sex of the recruit into account in such a situation.

In constructing Figure 2, the relationship between criterion and test was taken to be similar for the two groups, only the means of the test scores being represented as different. In Figure 3, on the other hand, a more complex difference is illustrated. In this figure the regression of criterion performance on test score is pictured as different for women than for men. Both the slopes of the regression lines and the errors of estimate are also taken as different. Evidence supporting this model as being correct would again point to the desirability of different procedures for women than for men. However, this situation is more complicated than that represented in Figure 2. Women do more poorly than men on the test, but on the average do as well on the job, as was also the case in Figure 2. However, in the representation in Figure 3, a given increase in test score is associated with a greater jump in criterion performance for women than for men. (For a score of X points, the most likely criterion performance is W for women and M for men. For a score of Y, these performances are W' and M', respectively. Note that the increase from W to W' is greater than the increase from M to M'.)

If only this single predictor is involved in the acceptance or rejection of a recruit for the particular assignment, again a different cutting score for women than men would be a simple means of handling the situation. However, if scores on the test are to be combined with other scores, then the model represented in Figure 3 will require more complex treatment than will the model in Figure 2. In particular, if linear composites are to be obtained through use of multiple regression equations, separate multipliers or regression weights will be needed for men and women for a given test.

The fourth figure illustrates still another possibility for which evidence might come to light in a survey of the literature. Perhaps there exist other predictive measures for each of which the scatterplot of criterion performance on test score will be the same for women as for men. If such measures could be found, they would have distinct advantages over present measures in terms of ease of interpretation and application.

Having discussed several possible models which available evidence might fit, it is now appropriate to turn to the review proper to see what data have in fact been published, and how these data might influence the course of further development of predictive measures to be used both with men and women.

B. Organization

At first it was hoped that this report might be devoted largely to the review of published articles reporting studies in which both men and women

- (a) had been tested with the same tests (the tests being similar to Navy tests),
- (b) had worked side by side on the same jobs (the jobs being similar to Navy billets), and
- (c) had been evaluated with the same measures and techniques for assessing performance.

In reading the next section of the review it will be observed that only a few reports could be found which fit this description even loosely.

Although many factory jobs are performed by both men and women, there seemingly has been little interest in a comparative analysis of the relation between on-the-job performance and selection test scores for men and women. This is true, at least, of published research. The experimental populations involved in most reported studies consist entirely of members of one sex. In many cases the selection measures utilized are ones that have been standardized on populations of one sex. Very often the same cutting scores are used quite uncritically for selecting workers of both sexes.

Rather than confine the attention of the report to the few studies which fit the specifications listed above, the scope of the report was expanded to include studies describing situations in which women had been selected for employment on jobs similar to billets to which women might be assigned in the Navy, both (1) with tests similar to those in the Navy basic test battery and (2) with other tests, even when comparable data for men were not presented. In considering the appropriateness for use with women of present Navy tests, one quite properly can raise the questions of whether tests like the ones being used do predict success on the job, and of whether there are other tests which might be preferable from a predictive viewpoint. Furthermore, a presentation of what has been published along these lines may point out the gaps which other research workers may then fill.

Another extension of the coverage was the inclusion of a review of some reports of sex differences on tests similar to tests in the Navy basic battery. Reported sex differences, of course, must be regarded as evidence of need for further information. The pertinence of such reports here is that they call attention to the need for additional data on these Navy-like tests so that the appropriate way to employ and interpret scores on such tests may become clear.

Contents of Later Sections of the Report. In Section II brief descriptions of Navy tests used with enlisted women are presented. Sections III and IV present reviews of two major contributions to the literature on appropriateness of selection and classification measures used with both men and women.

Sections V and VI contain reviews of studies in which women have been selected for jobs similar to those in the Navy and the effectiveness of the tests has been ascertained for predicting success on the job. The studies covered herein differ from those reviewed in Sections III and IV in that no comparable data are presented for men.

Section VII contains a review of reported studies in which sex differences have been observed on tests similar to those in the Navy basic test battery. Studies included in this section did not present follow-up validity data on the tests.

The final section, Section VIII, attempts to assess the significance of the literature surveyed, from the standpoint of implications for the appropriateness of selection and classification tests for use with enlisted women in the Navy.

With the exception of the British work and of the U. S. Air Force studies reported in the following two sections, most of the studies reviewed have been modest in scope, carried out by a single research worker or small group of workers over a short period of time. The aims have been practical, resulting in attempts to solve pressing problems of the moment. Consequently, the studies vary greatly in scope, method, choice of criteria, and subjects. Comparisons between them tend to be difficult as well as dangerous. Research designs vary in quality, and although techniques have improved in recent years, authors frequently neglected to describe fully the methods that were used.

II. Navy Tests Administered to Enlisted Women

Since a major focus of attention in this report is the tests used by the Navy for the selection and classification of enlisted women, it is quite appropriate to present brief descriptions of these tests. As findings from various military and industrial situations are presented, the reader may then be better able to note to what extent the results cited are for tests similar to Navy tests or are for tests somewhat different from these.

There are four tests in Form 5 of the Navy Basic Test Battery. These are (1) the General Classification Test, (2) the Arithmetic Test, (3) the Mechanical Test, and (4) the Clerical Aptitude Test. Each test is briefly described below. (The items provided as illustrations are similar to those in the tests but of course are not actual operational items.)

1. General Classification Test. This is a test composed of verbal analogies and sentence completion items. Items resemble the following:

PAGE is to book as TREE is to

(1) lumber (2) forest (3) paper (4) farm

If the radar indicates the approach of enemy aircraft,
a general will be sounded.

(1) barrage (2) control (3) observation
(4) battle (5) warning

2. Arithmetic Test. Items in Part I of this test call for the addition, subtraction, multiplication, and division of integers and fractions. Some items also involve the use of decimals and percentages.

The second part of the test is a measure of arithmetic reasoning. Items are of the following type:

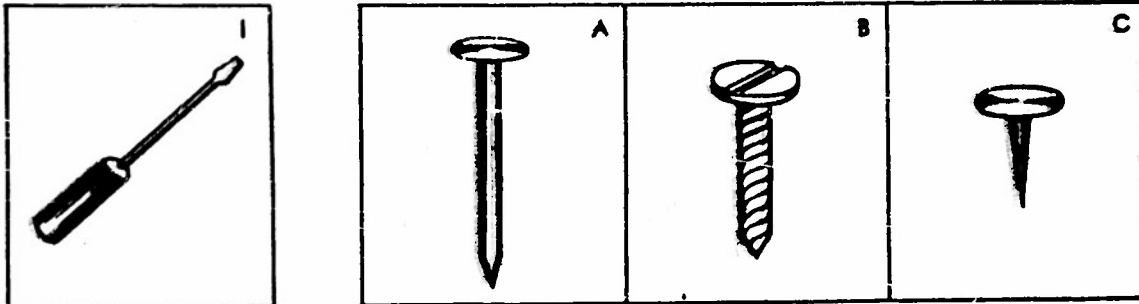
If a man walks a mile in 20 minutes, how many miles will
he walk in two hours at this pace?

(A) 2 (B) 3 (C) 6 (D) 8 (E) 10

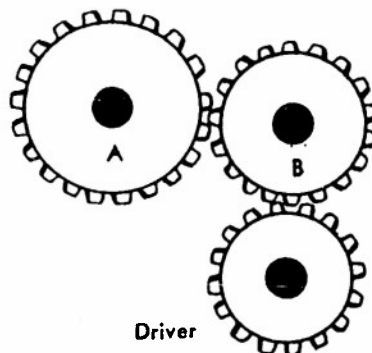
3. Mechanical Test. This test consists of two parts. The first part is a test of mechanical and electrical knowledge; the second, a test of mechanical comprehension. Both parts are in pictorial form.

In Part I each item involves four pictured objects. These may be tools, materials used in construction, electrical devices and components,

and the like. The examinee is called upon to state which one of three of the pictured objects is associated most closely with the first or stimulus object. Items resemble the following:



The mechanical comprehension part of the test is similar to Bennett's Test of Mechanical Comprehension (19). An illustration of the item type is given below:



Which of the gears turn in a direction opposite to that of the driver?

- (A) Gear A
- (B) Gear B
- (C) Both gears

4. Clerical Aptitude Test. Form 5A of this test consists of two parts, name checking and number checking. Content of this Navy test is similar to that of the Minnesota Clerical Test (5). The National Institute for Industrial Psychology (NIIP) Group Test 20 (Checking) (131) is similar to the second part of the test. Another test used in industrial situations, the Hay Number Perception Test also resembles part of the Navy test.

In the name checking part, the examinee is called upon to compare names presented in two columns and, as rapidly as he can, to

indicate whether these names are the same or different. An illustration is

	S	D	
Pioneer Forester Company	---	---	Pioneer Forestry Company
	S	D	
Benson's Refinery Works	---	---	Benson's Rebindery Works

The number checking part is arranged in a similar manner. An illustration of the item type is the following:

	S	D	
473926	---	---	473926
	S	D	
3200528	---	---	320528

III. Selection and Classification of Women in the British Armed Forces

Studies reporting the occupational selection of women with comparable data for men have arisen largely out of situations in which the problems are similar to those faced by the Navy. Reports of the selection of women in the British forces and in the U. S. Air Force were found to provide the most valuable sources of information. This section will present a review of British reports, and the following section will be concerned with the U. S. Air Force studies.

Reports on the selection procedures used by the British during World War II indicate that before the end of the war over 200,000 women had been selected for a total of 106 jobs in the Auxiliary Territorial Service (ATS). Routine written reports of success in training apparently were available for 39,000 of these women who entered the ATS between October 1942 and September 1943. Special reports were prepared on all auxiliaries failing in the training courses. In addition, some measure of on-the-job proficiency was obtained for a total of 5000 auxiliaries who were followed up in 27 particular trades and employments by means of visiting, collecting examination results, and holding rating conferences on proficiency.

In view of the large number of women selected, the variety of jobs performed, and the opportunity for follow-ups beyond the training period to actual performance on the job, it was at first expected that this material would provide answers to many of the questions that led to the preparation of this review. This did not prove to be so, for, as Vernon (183) has pointed out, the exigencies of war created an extreme shortage of trained staff, of Personnel Selection Officers familiar with statistical methods, and of automatic calculating machines, and, further, prevented any planned experimental approach. As a result, comparisons of the British data with data produced in this country tend to be difficult.

In the British reports published since the war, criteria have seldom been explicitly stated. This may be due in part to the fact that they seem to have varied from situation to situation. During wartime, characteristics of score distributions were reported in terms of medians

and percentiles because these statistics were considered to be more easily interpreted by the inexperienced personnel who used them, even though, as Vernon (183) noted, these distributions were frequently skewed. Problems are further complicated for the American reader as job descriptions were not supplied, and there is reason to believe that similar titles did not always imply similar responsibilities.

In some cases the tests in use with women differed from those in use with men, even at the outset. In other cases, identical tests were weighted differently for women than for men. Validity coefficients were reported for job groupings rather than for single jobs, and frequently these groupings do not correspond to job groupings in U. S. women's forces for which corresponding data are available.

Selection for jobs in the ATS (as in other branches of the British services) was not based entirely on test scores. Interviews and a "Qualification Form" (biographical inventory) supplemented test information. Vernon (184) stated that no evidence was shown that objective testing could entirely replace the interview. Vernon and Parry (186, p. 287) indicated a belief that the interview was "essential on the grounds of flexibility and humanity, in spite of its inaccuracy."

The test battery used for selection and classification varied from one branch of the service to another. While no attempt will be made to describe all the tests in detail, a brief description of measures referred to in this report is given below. For a fuller account the reader should consult Vernon and Parry (186).

The Progressive Matrices Test was adopted as the primary general intelligence test in the Royal Navy, Army, and ATS in 1941. It is a non-verbal intelligence test. Each item requires the subject to induce relationships among geometric figures in a matrix in order to select a figure which completes the pattern (186, p. 234).

The Instructions Test was a clerical test which required the examinee to perform the operations of checking, filing, classifying, and coding printed information in rapid rotation (186, p. 222).

The Spelling Test used in the ATS required the examinee to choose the one correct spelling of a word from a list of six different spellings. A synonym at the beginning of each line helped the examinee to identify the word (186, p. 227).

The Arithmetic and Mathematics Tests used typically contained two parts: the first consisting of straightforward addition, subtraction, multiplication, and division problems, and the second including thirty forty brief problems (186, p. 228). These tests were considerably simplified for use in the ATS.

The Spatial Test used most widely in the British Navy, Army and ATS was the NIIP Squares Test. This consisted of a series of fifty figures in each of which the examinee was required to draw a dividing line such that the two pieces so formed would, if turned around, make a square (186, p. 236).

Bennett's Test of Mechanical Comprehension was one of several mechanical tests used in the Navy, Army and ATS. In the ATS a Practical Problems Test roughly comparable to Form W1 of the Bennett test was also introduced (186, p. 241). (Form W1 is a form designed for women.)

A Meccano Assembly Test was used with great success in the ATS as a supplement to the Bennett (186, p. 242). (Meccano is a construction set consisting of punched strips, wheels, gears, pulleys, and plates, out of which it is possible to assemble operating models of many mechanical devices.) In the Army a similar but more difficult assembly test was used.

A considerable degree of success was reported for the selection procedures used in the ATS (186, p. 48). Wickham (191) provided correlations between scores on six tests of the ATS battery and the criterion of success in training for 27 ATS jobs. Beta weights and multiple correlations are also cited. A summary of these data is given in Table 2. On-the-job validity data are presented in Table 3 for twelve job groupings in the ATS. The number of cases on which correlations were based varied from 30 in some samples to 1128 in the largest sample; the median sample size was 106 (186, p. 210; 191, Table 4). In another report

Mercer (128) observed that multiple correlations (corrected, presumably for restriction of range) for the test battery with success in 27 particular trades varied from .41 to .96 with a clustering around .70.

Some information which might be used as a gauge of the relative effectiveness of test batteries used in the ATS and other British services was given by Vernon and Parry (186, p. 212). The results for the principal tests are summarized in Tables 4 and 5, which present not only the median validity coefficients in all comparable Naval and ATS studies, but also provide a notion of the range of the observed coefficients.

Mercer (128) indicated that, of the 39,000 auxiliaries passing through selection procedures and undergoing training for specific ATS trades and employment, over 94 per cent were successful in training. Mercer also provided figures, summarized in Table 6, showing the decrease in failure rates following the introduction of a selection program. Vernon and Parry (186, p. 120) reported training failure rates of 10,000 Army tradesmen selected by four different methods in 1942 and trained simultaneously. These rates varied from 27.1 per cent for men called up by the Ministry of Labour as semi-qualified tradesmen to 17 per cent for men nominated by the commanding officer or other technical officer, to 16.6 per cent for men nominated at their own request, and 8.7 per cent for men selected by Personnel Selection Officers. Similarly, Vernon and Parry (186, p. 121) indicated that failure rates for men trained as Fleet Air Arm mechanics and fitters fell from an average of 14.7 per cent to 4.7 per cent with the introduction of the new selection methods.

From the findings cited above, it seems reasonable to assert that the effectiveness of the selection procedures in use in the ATS was comparable to that of the procedures in use in the Army and the Royal Navy insofar as it is safe to assume that criteria of success and failure remained stable with passage of time and from service to service.

The rather high validities reported for the tests in the ATS battery may be accounted for in part by the fact that no attempt was made to use with women the same tests and weights used for selecting men for similar jobs. Tests were standardized, and weights were established

Table 2

Zero-order Correlation Coefficients, Regression Weights, and Multiple Correlation Coefficients Corrected for Homogeneity for the Basic Tests as Predictors of Training Success for 22 Job Groupings in the ATS

Adapted from Wickham (191, p. 160)

	Spatial (squares)		Paper Mechanical (Bennett)		Progressive Matrices (1938)		Arith- metic (ATS)		Instruc- tions (Clerical)		Spelling (ATS)		Malt. corr.*
	r	β	r	β	r	β	r	β	r	β	r	β	
N	482												.76
Clerks, General Duties	.31	-.03	.31	.06	.58	.22	.56	.07	.61	.10	.68	.32	.75
Clerks, R.A. Ordnance Corps	.19	-.01	.22	.14	.26	-.02	.26	--	.37	.20	.30	.08	.41
Clerks, R.A. Service Corps	.48	.16	.37	.11	.57	.16	.60	.30	.56	.09	.55	.14	.72
Clerks, Surveyors	.28	.02	.23	.05	.42	-.06	.46	-.02	.62	.46	.55	.14	.66
Storewomen, Tech.	.31	.10	.21	.05	.37	.09	.42	.22	--	--	.39	.17	.48
Cooks	.112	--	.25	.08	.44	-.08	.64	.39	.65	.43	--	--	.71
Telephonists	.39	.05	.18	-.11	.55	.01	.52	.24	.51	.19	.38	.03	.59
Operators, spec. wireless	.32		.22	-.09	.68	.41	--	--	.66	.20	.65	.28	.78
Operators, teleprinter													
Operators, wireless and line	.32	.07	.31	.14	.43	.05	.51	.23	.44	-.01	.49	.18	.62
Operators, switchboard	70	--	.30	.11	.36	--	--	--	.46	--	.65	.28	.74
Operators, cipher (med. grade)													
Operators, cipher (high grade)	.45	.05	.31	.03	.64	.23	.69	.32	.65	.11	.63	.10	.79
Operational numbers:	81	.41	.35	.15	.50	--	.55	--	.82	.64	.54	.13	.80
Height-takers	.185	.46	.22	--	.44	.11	.45	.21	.56	.40	--	--	.59
Spotters	96	.51	.22	-.06	.54	.13	.48	.08	.59	.35	--	--	.66
Predictor operators	265	--	.32	.16	.46	.19	.51	.37	--	--	--	--	.57
Radar operators	101	.22	.23	--	.53	.08	.69	.45	.65	.25	.55	.01	.77
Kine-Theodolite operators	59	.23	.19		.52	.20	.44	.22	.52	.05	.37	--	.67
Drivers	1128	--	.30	.11	.27	-.05	.25	-.09	.41	.26	.27	-.03	.49
Fitters, motor vehicle	140	.11	.40	.36	.23	.02	.18	.14	.03	--	.23		.43
Fitters, general	83	.44	.40	.21	.37	.02	.25	-.19	.31	.20	--		.65
Electricians	72	.51	.25	--	.64	.23	.56	.07	.56	--	N	N	.73

*Multiple R includes some additional tests.

Note: Where more than one criterion was used the correlations were averaged. A blank indicates results are unknown; a dash (--) that they were known, but negligible. N indicates results were known for limited numbers -- but were not comparable with those for the rest of the battery.

Table 3

Regression Weights and Multiple Correlation Coefficients (Corrected for Homogeneity) for the Basic Tests
as Predictors of On-the-job Success for Twelve Job Groupings in the ATS.

Adapted from Wickham (191, p. 165)

	N	Beta Coefficients						Multiple Correlations*
		Spatial (Squares)	Paper Mechanical (Bennett)	Progressive Matrices (1938)	Arith- metic (ATS)	Instructions (Clerical)	Spelling (ATS)	
Clerks R.A.Ordinance Corps	198	--	--	.145	.09	.097	--	.29
Storewomen (tech.)	340	--	--	.030	.20	.161	--	.35
Telephonists	250	.03	.118	-.097	-.02	.409	--	.40
Spec. Wireless oper.	500	.09	--	--	.19	.125	N	.48
Height-takers	250	-.02	.138	.136	.22	.018	.10	.46
Spotters	140	.07	.074	.121	-.06	.108	.11	.33
Predictor operators	100	.00	.037	.239	.10	.147	-.07	.40
Radar operators	250	.05	.016	.203	.10	.131	.00	.50
Drivers	200	--	--	--	-.03	.094	-.05	.16
Orderlies	91	N	--	--	.15	.135	N	.24
Plotters	55	-.15	-.012	.458	.15	-.112	.31	.54

*Multiple R includes confirmatory tests and educational standard.

Note: dash (--) means known but negligible

N indicates results known for limited number -- but not comparable with rest of the battery.

Table 4

Descriptive Statistics of the Distributions of Raw Validity Coefficients
and Multiple R for Standard Naval Selection Tests

Adapted from Vernon and Parry (186, p. 212)

<u>Statistic</u>	<u>Tests</u>					<u>Multiple R</u>	
	<u>Pro- gressive Matrices</u>	<u>Shipley Abstrac- tion</u>	<u>Bennett Mech. Comp.</u>	<u>Mathe- matics</u>	<u>Squares</u>	<u>T2*</u>	<u>(uncor- rected)</u>
90th per- centile	.45	.49	.44	.57	.38	.57	.70
50th per- centile	.28	.30	.28	.35	.22	.40	.47
10th per- centile	.10	.11	.13	.17	.05	.20	.32

*T2 refers to Total Score (with Test 1 doubled) on the Royal Navy's standard test battery. The battery consisted of four group tests:
1. Modified Shipley Abstraction, 2. Modified Bennett Mechanical Comprehension, 3a. Arithmetic, 3b. Mathematics, and 4. Squares Test of Spatial Judgment.

Table 5

Descriptive Statistics of the Distributions of Validity Coefficients
Corrected for Multivariate Selectivity and of Multiple Correlation
Coefficients for Basic Selection Tests in the ATS

Adapted from Vernon and Parry (186, p. 212)

Statistic	Tests					Multiple R	
	<u>Pro- gressive Matrices</u>	<u>Bennett Mech. Comp.</u>	<u>Arith- metic</u>	<u>Squares</u>	<u>Cler- ical</u>	<u>uncor- rected</u>	<u>cor- rected</u>
90th per- centile	.65	.41	.69	.51	.66	.69	.84
50th per- centile	.49	.30	.51	.40	.56	.47	.65
10th per- centile	.27	.19	.26	.20	.35	.35	.50

Table 6

Failure Rates in Four ATS Job Groups Prior to and Following
Introduction of Selection Procedures

Adapted from Mercer (128, p. 196)

Job	Pre-Introduction		Post-Introduction	
	Number Selected	Failure Rate	Number Selected	Failure Rate
Clerks	128	11%	592	4%
Drivers	124	30%	1004	14%
Special Operators	420	64%	130	8%
Operators: Wireless and Line	217	7%	187	0.5%

for use with women, through studies of ATS personnel. Although this procedure may indicate excellent judgment on the part of those responsible, it complicates the task of analysis undertaken here. A simpler arithmetic test was devised for the ATS than was used with men, since women seemed to have difficulty in handling decimals. In selecting mechanics, a modified Bennett test was supplemented by a Meccano Assembly Test. Vernon (185) reported that this combination of the Bennett and the assembly test showed greater differentiation in selecting women for mechanical jobs than did the Bennett, used alone, in selecting men for similar jobs. He concluded elsewhere (184) that tests of mechanical comprehension were acceptable for use with women and adolescents who had little previous experience, but that for adult males, straightforward information tests were more promising.

Vernon further stated that mathematics tests seemed to give higher correlations with proficiency at the end of training for mechanical jobs than did any of the mechanical tests. In the absence of any information about methods of selecting the samples upon which these conclusions were based, these assertions should perhaps be qualified. As Vernon himself pointed out in other discussions, validity figures for any particular selection variable suffer to the extent to which personnel have been selected on that particular variable. As a result, some other test, playing no role in the original selection, may achieve an apparently closer relationship to the criterion variable than is observed for the variable on which selection has occurred. Vernon and Parry (186, p. 213) reported that despite the fact that for some jobs all of the tests used yielded high validity coefficients, and for other jobs all coefficients were low, "the relative validities of the different tests were remarkably uniform."

According to Vernon and Parry, the tests were mainly used to distribute the available supplies of high quality personnel among the different branches according to their needs. Differentiation between jobs was based more on interests and interview judgments than on test scores. (186, p. 215). Vernon and Parry speculated upon the reasons for the tendency of the "verbal: educational" tests to correlate in general more

highly with measures of proficiency than any of the other tests administered. They offered several possible explanations for these findings (186, p. 215-216): the high reliability of verbal and clerical tests; the high g-saturations of these tests; the fact that these tests possibly involved measures of certain personality or temperamental qualities important to vocational success; the extreme heterogeneity of the samples in respect to g; and tendency of the jobs to be more varied than most civilian employments. These authors also suggested that if more objective measures had been available as training criteria, and if training criteria had been supplemented with assessments of operational efficiency, the "verbal: educational" tests might have yielded lower correlations, and specialized tests higher correlations with the criteria. Finally, the authors considered that had there been more time available, tests more successful than these might have been found.

In order to check on the possibility that the choice of training grades as criteria led to the high validities of the "verbal: educational" tests, Vernon and Parry presented some findings based on operational follow-ups. In one study, assessments of efficiency during fighting in Italy were collected for 200 Royal Marine signallers (186, p. 216). Naval selection tests were administered to the men after they returned to Britain. The T2 scores on the standard Royal Navy Battery (see footnote, Table 4) correlated .62, and component tests between .46 and .49, with measures of efficiency. It can be seen by referring to Table 4 that the validities of T2 were as high or higher when efficiency during fighting served as the criterion as when training grades were used. In another study (186, p. 217), over six hundred trainees for anti-aircraft duties in the ATS were followed up, and later some thirteen hundred women were assessed for efficiency after serving two or more years. The average validity coefficients corrected for selectivity are shown for both groups in Table 7.

Brown and Ghiselli (36) compared for a number of tests given to several job groups the validities against training criteria with on-the-job validities. They obtained 127 pairs of such validities from research reported in the literature. These authors differed from Vernon and Parry

(186) claiming that the relationship between validities obtained in training situations and on the job tends to be low; Brown and Ghiselli did state, however, that the relationship between the two kinds of validity figure was higher for intelligence and clerical tests than for tests of other kinds. They also indicated that the relationship between the two types of validity was highest for studies involving "manipulative and observational" jobs.

Table 7

Mean Validities (corrected for selectivity) of A.T.S. Selection Tests in Several Anti-aircraft Jobs at Different Stages

Adapted from Vernon and Parry (186, p. 217)

	Tests						Multiple R
	Pro- gressive Matrices	Bennett Mech. Comp.	Arith- metic	Squares	Cler- ical	Spell- ing	
Training stage (N = 600+)	.47	.24	.53	.11	.42	.06	.65
Operational stage (N = 1300+)	.35	.21	.30	.25	.37	.31	.43

IV. Performance of Women in Air Force Technical Schools

An opportunity for making comparisons between the performance of men and women on the same tests and in the same training situations has been provided by some Air Force research. A study by Howard and Pickrel (97) compared samples of women in the Air Force (WAF) with samples of male airmen in seven Air Force technical schools. Aptitude indices from the Airman Classification Battery and final course grades in these schools were studied for all samples (see Table 8). The seven courses provided training for clerk-typists, radar operators, radio mechanics (general), radio operators, supply technicians, teletype operators, and weather observers.

Howard and Pickrel were primarily interested in investigating the relative validities of the battery for men and women; that is, they wished to determine whether the battery would predict the success of the women sufficiently well to warrant its use in the selection of women for Air Force technical schools. The authors discovered that the validities of the aptitude indices were generally lower for WAF samples than for samples of male airmen (see Table 8), but concluded that the differences were slight and that the aptitude indices from the Airman Classification Battery predicted grades sufficiently well to justify their continued use in selecting both WAF and male airmen for technical schools.

It is interesting to note, however, that in each of the seven schools the mean final grade for the women exceeded that for the men. This is true, for instance, for radio mechanics, radio operators, and weather observers, in spite of the fact that the mean score on the selection variable was lower in these cases for the women (see Table 8). These facts bear important implications in that the exact nature of the relationship between aptitude scores and course grades is different for women than for men; that is, the final course grade predicted from any given score on the Airman Classification Battery will differ for WAF from that predicted for male airmen.

Table 8

Mean Scores and Standard Deviations on the Selection Variables and Final Course Grades, and Correlation Coefficients between Test Scores and Grades for White Male (WM) and WAF Samples from Seven Air Force Technical Schools

Adapted from Howard and Pickrel (97, p. 11-17)

Technical Course	N		Mean Score		Standard Deviation		Mean		r	
	WM	WAF	WM	WAF	WM	WAF	WM	WAF	WM	WAF
Clerk Typist	1965	205	6.63	6.97	1.33	1.10	3.67	4.13	.63	.61
Radar Operator	518	59	6.67	6.98	1.45	1.07	3.58	3.70	.57	.53
Radio Mechanic (Gen.)	513	56	7.46	6.11	1.29	1.26	3.40	3.65	.66	.52
Radio Operator	321	72	7.44	7.29	1.20	1.36	3.44	3.81	.49	.46
Supply Technician	414	225	6.15	6.84	1.42	1.20	3.15	3.47	.64	.67
Teletype Operator	599	214	6.10	6.44	1.45	1.32	3.79	4.08	.52	.44
Weather Observer	266	96	7.43	7.29	1.38	.93	3.18	3.42	.55	.30

Mary Agnes Gordon (78) carried the analysis one step further in an attempt to determine the feasibility of using the same minimum qualifying scores for women as for men in selecting students for Air Force technical schools. By studying the regression of final school grades on scores on the selection measures, Gordon provided the basis for a somewhat more sensitive analysis of the relation between the selection variables and the criterion than had hitherto been attempted.

Gordon reported that the regressions of final grades on aptitude scores for the WAF group differed sufficiently in every technical school sample from the regression for the white male group to justify the conclusion that "the same aptitude index will not predict the same final school grade for these groups" (78, p. 6). The regressions for WAF differed from those for white males at the .01 level in all but one sample, that for the Radar Operator School. Significant differences in errors of estimate occurred in the second Clerk-Typist and in the Photo Laboratory Technician samples. A significant difference in slopes was found in the Supply Technician sample. Significant differences in intercepts occurred in the remaining samples. All significant differences in intercepts favored the WAF (see Table 9).

The difference was most striking in the case of the Radio Mechanic School. In terms of a five-point grade scale with 2.5 the passing grade and 5.0 the maximum grade, the grade predicted from an aptitude index of five was .42 grade points higher for WAF than for white males. The difference was .30 grade points in the Radio Operator School, .32 grade points in the first Clerk-Typist School, and .24 grade points in the second Teletype Operator School. Even larger differences in favor of WAF personnel occurred in the grades predicted from an aptitude index of nine. The Photo Laboratory Technician School was the only one in which the difference was slightly in favor of white males.

At the .05 level of significance the slope of the regression line for the WAF group differed from that for the white male group in three samples. The difference in slopes is apparent in the differences in predicted grades for aptitude indexes of five and nine presented in Table

Table 9

Differences in Final School Grades Predicted from Minimum and Maximum
Aptitude Indexes of the White WAF Group from the White Male Group

Adapted from Gordon (78, p. 6)

<u>Technical School</u>	<u>Aptitude Index</u>	<u>Difference between WAF and white male predicted final school grade in grade units*</u>
Clerk-Typist	5	.32
	9	.56
Supply Technician	5	.14
	9	.32
Teletype Operator	5	.11
	9	.22
Radio Operator	5	.30
	9	.39
Radar Operator	5	.09
	9	.09
Radio Mechanic	5	.42
	9	.46
Weather Observer	5	.05
	9	.40
Clerk-Typist (Second sample)	5	.24
	9	.38
Teletype Operator (Second sample)	5	.24
	9	.21
Photo Laboratory Technician	5	-.05
	9	-.07

*A positive figure indicates that the predicted grade for WAF was higher than that predicted for white male airmen. A negative figure indicates that it was lower.

9. The respective differences are .32 and .56 for the first Clerk-Typist sample, .14 and .32 for Supply Technician, and .05 and .40 for Weather Observer. In each case the intercept (at an arbitrary stanine of 5) is higher for the WAF group.

Gordon pointed out that further examination of the data showed there was no combination of predictors for these or other samples that would equalize the group differences in regressions. When the regression of final grades on Biographical Inventory scores was similarly studied, a significant difference in intercepts was found for all samples.

Gordon concluded by recommending that, when a minimum aptitude index of five is required of white males, the following minimum indexes should be used for WAF:

- (1) An aptitude index of four for schools in the Clerical Cluster.
- (2) An aptitude index of four for schools in the Radio Operator Cluster.
- (3) An aptitude index of five for schools in the Technician Specialty Cluster.
- (4) An aptitude index of two for schools in the Mechanical Cluster. As results were based on only one school, Gordon offered this recommendation only tentatively.

Gordon added in a footnote, however, that "If different minimum aptitude indexes are to be recommended, it is important to determine regions of significant differences between sample regression lines which are non-parallel" (78, p.8).

Several factors considered by Gordon in her attempt to explain why WAF personnel earned higher course grades than white males having the same aptitude index were the following: (1) possible cultural bias of tests in the Airman Classification Battery; (2) greater motivation on the part of the WAF; (3) bias in the final school grade in favor of WAF because of neatness, superior spelling, a greater ability to express themselves; or (4) a general tendency of women to overachieve in school.

The work of Gordon and that of Howard and Pickrel have provided the most pertinent and hence valuable material in terms of the problems faced by the Navy. The studies are, in fact, the only studies which give some clear indication of the relative performance of men and women on Navy-like tests, and on Navy-like jobs. Here, then, is a strong suggestion that predictions from scores on the Navy Basic Test Battery may differ in meaning, when women are measured, from predictions when men are measured, although with some adjustment in minimum cutting points the scores may yet be extremely useful.

The final test of the advisability of lowering the minimum qualifying scores for technical schools in the case of women must rest, of course, upon information derived from measurement of on-the-job proficiency. Such information is not as yet available in the literature, although in this connection Wickham (191, p. 169) noted a parallelism in the British studies between training validities and on-the-job validities. "The results found for follow-up in the working units practically duplicated those found for training."

V. Civilian Studies of the Selection of Women for Navy-like
Jobs Using Tests like Those in the Navy Basic Test Battery

A. Clerical Workers

The Navy utilizes a considerable number of enlisted women in clerical positions. Reports on the selection of women for clerical work form the largest single group of research studies in the literature which may be considered relevant to this survey. Tests similar to some of the Navy's tests, such as the Minnesota Clerical Test, frequently have been used. Unfortunately, the literature afforded little by way of comparisons between the selection of male clerical workers and the selection of female clerical workers. Most clerical workers are women, and most studies of selection of clerical work have been based on women. Occasionally men formed part of a group that was studied, but usually no break-down of the results by sex was attempted. The policy has seemingly been, rather, to insure greater experimental control by omitting data from one sex or the other, especially in cases similar to that reported by one author (103) in which significant sex differences in performance were noted.

The discussion, therefore, will be limited mainly to the selection of women for clerical work. In general, tests found to be reasonably successful in selecting clerical workers were of three types: tests of general intelligence, tests of clerical aptitude, and measures of performance on work samples which involved typing, filing, etc. In several cases a battery combining tests of two or more of these types was used. It should perhaps be added that in none of the multiple correlation studies on the selection of clerical workers reported here was there evidence of cross-validation nor was evidence cited to demonstrate that cutting scores derived from one sample would serve as successfully to select good workers in another sample.

1. The Use of Test Batteries in Selecting Clerical Workers. One of the earliest studies reporting the use of a battery in selecting clerical workers was conducted in 1921 by Bills (24, 25). She concluded from her study of 139 applicants for courses in stenography and comptometer operation that a battery of tests was more effective than any

single test in picking successes and eliminating failures. Her battery included a test of general intelligence adapted from the Army Alpha, and a test of aptitude for typing and stenography. Of single tests, she concluded that a test of general intelligence was most effective for eliminating persons likely to fail and that a test of special ability, e.g., stenography and typewriting, was most effective in picking successes.

For selecting bank machine bookkeepers, Hay (88) reported the successful use of two batteries similar to Bills' in that they utilized clerical aptitude and general intelligence measures as selectors. One battery consisted of Alpha Number Series, Minnesota Numbers, and Fryer Name Finding, and the other battery contained the Otis Self-Administering Test of Mental Ability (Form B), Minnesota Names, and Minnesota Numbers.

Using the Wherry-Doolittle method for selecting a test battery, Holmes (96) found that a combination of the Wonderlic Personnel Test with a typing test yielded a multiple correlation of .48 with criterion ratings for 88 secretaries; a combination of a cancellation test from the State Farm Personnel Survey and a typing test yielded a multiple correlation of .44 with criterion ratings for 56 typists; and a combination of the Wonderlic Personnel Test with a general Clerical Ability Test from the State Farm Personnel Survey yielded a multiple correlation of .52 with criterion ratings for 107 clerk-typists.

Hay (89) reported a multiple correlation of .38 between scores on a combination of the Minnesota Clerical Test, a number series completion test, a name finding test, and supervisor's ratings for a group of 82 key-punch operators. In a later article (90), Hay reported the predictive efficiency for a battery of five tests applied to 82 key-punch operators. The tests used were: Wonderlic Personnel Test, Minnesota Clerical (Numbers and Names), Hay Number Series A, and Hay Name Finding. Five different combinations of cutting scores were successful to the extent of selecting a group of which 85 per cent or more of the workers were rated "good" by supervisors. The proportion of workers rated "good" in the group falling below the cutting score varied from 15 per cent to 40 per cent in the five examples, and in two of the five cases more than 50 per cent of the workers rated "good" fell below the cutting score. No cross validation was reported.

2. Reported Validities for Single Tests Used in Selecting Clerical Workers. Reports of the relations between various criteria of job success for clerical workers and scores on tests, taken singly, are available

from some of the studies referred to above. Tests which are similar to the Navy's tests have been emphasized in the discussion below.

Validity coefficients for the Minnesota Vocational Test for Clerical Workers have been reported as follows: .30 with supervisors' ratings for 82 key-punch operators by Hay (89); Minnesota Numbers .62 and Minnesota Names .54 with measures of production for 27 typists by Blakemore (28). Although he provided no validity coefficients, Barrett (11) indicated that the Minnesota test differentiated between good and poor students in a typewriting course. For a group of 39 machine bookkeepers Hay (88) reported a correlation with production of .51 for Minnesota Numbers, and of .47 for Minnesota Names. In a study of 120 bank clerks and typists Seashore (158) reported validity coefficients for Minnesota Numbers of .29 with speed, .36 with accuracy, and .30 with "ability to grasp new ideas." Minnesota Names correlated .38, .39, and .45 with these criteria, and the total score gave corresponding coefficients of .38, .42, and .43.

In one study Hay (89) reported the following correlations with supervisory ratings: .25 for Number Series Completion, .26 for Name Finding, and .30 for the Minnesota Clerical. The subjects were a group of 82 key-punch operators, presumably female.

In studying 27 female typists for whom the criterion was a measure of production, Blakemore (28) found that the Hay Number Perception Test and the Minnesota Clerical Test (Numbers) both had validity coefficients of .62. However, scores from the Hay Number Series test yielded no significant correlation in the same situation.

Holmes (95, 96) found that the general clerical sub-test of the State Farm Personnel Survey had the following validities against supervisors' ratings: .00 for a group of 88 secretaries, -.15 for a group of 56 typists, .42 for a group of 107 clerk-typists, .52 for a group of 100 supervisory personnel, and .49 for 50 interpretive personnel. Corresponding validities for the general information sub-test were .25, -.03, .15, .43, and .16 respectively. The cancellation test from the same battery yielded correlations of .08 for 88 secretaries, .33 for 56 typists, .11 for 107 clerk-typists, .62 for 100 supervisory personnel, .23 for 50 interpretive personnel, and .42 for a group of 50 key-punch and verifier operators.

In a study of office personnel, Giese (74) reported the following correlations between supervisors' ratings and various scores on the General Clerical Test: for the total score, .41 (N = 26); the clerical subscore, .34 (N = 38); the numerical subscore, .42 (N = 41); the verbal subscore, .48 (N = 26). A study of clerical workers cited in the manual of the General Clerical Test (144) reported correlations with performance ratings of .43 (N = 73) for the total score, .45 (N = 68) for the clerical subscore, .22 (N = 71) for the numerical subscore, and .42 (N = 70) for the verbal subscore.

Seashore (158) also reported validity coefficients for the General Clerical Test. For 120 clerks and typists these ranged from .56 when the criterion was ability to grasp new ideas as rated by supervisors, through .48 with ratings of speed, to .46 with ratings of accuracy. A biserial coefficient of .32 with grades in a training course was observed for a group of 116 clerical workers. In the latter study scores on the SRA Clerical Test gave a biserial correlation of .36 with letter grade in the training course.

The use of the American edition of the NIIP Clerical Test in the selection of students for college library work was reported by Oberheim (134, 135). Scores on the clerical test correlated significantly with the criterion ratings for men and women combined. In one sample the correlations were higher for men, and insignificant for women.

Although Oberheim (134) found that scores on the American Council on Education Psychological Examination correlated .42 with success in library work for men and women combined, she was reluctant to draw any conclusions about the relative value of the clerical test and the ACEPE in selection of the women because all correlations for the women were low. However, she continued to use a combination of the ACEPE, NIIP Clerical, and college grades as a predictor.

Holmes (95) found a correlation of .47 between scores on the Otis Self-Administering Test of Mental Ability and criterion ratings for a "skilled" group of office personnel. Hay (88) reported correlations of .56 between scores on the Otis SA (20' version) and production measures for 39 machine bookkeepers. In the previously mentioned report by

Seashore (158) of a study of 120 bank clerks and typists for whom the criteria were ratings made by supervisors, Otis scores correlated .17 with speed, .16 with accuracy, and .38 with ability to grasp new ideas.

Holmes (95,96) reported the following correlations between scores on the Wonderlic Personnel Test and criterion ratings: .33 for secretaries, .22 for typists, .36 for clerk typists, .51 and .49 for supervisory office personnel, and .56 for an "interpretive" group of office workers. Hay (90) reported a significant difference between mean scores on the Wonderlic for 53 key-punch operators rated "good" and 29 rated "poor." For 27 typists, Blakemore (28) found a validity coefficient of .32 on the Wonderlic test when a measure of production was the criterion. In studying 116 clerical trainees, Seashore (158) reported a biserial correlation of .35 between scores on the Wonderlic and letter grades in a training course.

Tiffin and Lawshe (176) reported higher scores on an Adaptability Test for 50 good clerical workers as opposed to 38 poorer workers. Rogers (149) used four subtests of the Woodworth-Wells Series, verb-object, number-checking, color-naming, and action-agent, in studying three groups. For one group of 77 students the criterion was mid-year grades in stenography, grammar, and typewriting; for the other two groups, consisting of clerical workers ($N = 38$ and 65), the criterion was a measure of production. Rogers found that these subtests correlated more highly with course grades than with measures of production, the coefficients ranging from .39 to .46 when course grades were the criterion, and from .13 to .39 when a measure of production was the criterion.

In a study of the effectiveness of grades on the Civil Service Examination for discriminating between various levels of card-punching efficiency, Marcus (125) found that a team of five subtests of the Woodworth-Wells Series correlated .45 with efficiency records based on speed and accuracy.

A general intelligence test adapted from the Army Alpha was found by Bills (25) to make correct predictions for 85 per cent of the applicants in courses in stenography and comptometer operation: those

who both failed on the test and failed on the course together with those who both passed the test and passed the course made up 85 per cent of the sample studied. Further computation by one of the present writers yielded a contingency coefficient for Bills' data in the neighborhood of .56 for a group of 67 subjects.

Several correlation coefficients between scores on subtests of the Army Alpha and production measures for 39 machine bookkeepers were reported by Hay (88) as follows: total score .51; number series .56; same-opposite .47; verbal .47; numerical .44; relationships .43; analogies .42; information .40; sentences .40; arithmetic .37; directions .32.

Kinney (103) reported biserial correlations of .63 between scores in addition and ratings for 77 female mail-order house clerical workers and .21 between scores in addition and ratings for 54 wholesale office workers. Seashore (158) reported correlations of .28, .25, and .39 between Alpha Number Series and ratings of speed, accuracy, and ability to grasp new ideas made by supervisors of 120 clerks and typists in a bank.

In spite of the large number of studies on the selection of clerical workers, the implications of the findings are by no means clear or unequivocal. No attempt has been made by the authors to relate the choice of criteria to the success of the tests. Criteria vary widely, from quite specific items such as measures of speed, accuracy, or a combination of both, to global ratings of the employee's general worth to the company. Nor has the influence of the level of responsibility held by the employee, undoubtedly an important variable, been related systematically to the selection measure except in a few cases such as that reported by Hay (90).

The reasonable degree of effectiveness of such tests as the Minnesota Clerical Test, the General Clerical Test, the SRA Clerical Test, and the NIIP Clerical Test has been amply demonstrated. Correlations with the criteria reported by authors quoted in this review tend to group around .42. Intelligence tests, such as the Otis SA, the Wonderlic Personnel

Test, the Woodworth-Wells Series, and selected subtests of the Army Alpha reported in this review yielded correlations with the various criteria averaging in the neighborhood of .40.

By surveying the literature and grouping reported validities according to job and test, Ghiselli and Brown (72) presented a more clearly organized picture of the status of tests in predicting the trainability and proficiency of clerical workers than had hitherto been available. They compiled considerable information about three types of clerical workers: general clerks, recording clerks, and computing clerks. While no reference was made to the sex of the subjects in their article, some of the conclusions drawn by Ghiselli and Brown with regard to clerical workers might be considered relevant to this review, for it is likely that the samples were made up largely of women.

On the basis of Ghiselli and Brown's analysis of available data, the tests having the consistently highest validities for all three types of clerical workers appeared to be arithmetic and number comparison tests. The mean of the reported validity coefficients for an arithmetic test was .43 for general clerks, .41 for recording clerks, and .35 for computing clerks. The mean reported validities for number comparison for these groups were .42, .29, and .33 respectively. For general clerks alone, intelligence, arithmetic, number comparison, and name comparison tests were the best tests and were about equal in validity. The mean of the reported validities for intelligence tests with this group was .42; for arithmetic, .43; for number comparison, .42; and for name comparison, .40. (The mean validity coefficients reported by Ghiselli and Brown were weighted means computed through Fisher's Z' transformation.)

Although Ghiselli and Brown were able to compile this information about three types of clerical work, they did not (or were unable to) consider the reports in the literature in terms of choice of criteria, level of job, or degree of responsibility held by the worker. It is possible that some information regarding the differential value of the test is lost when such an analysis is not carried out. For general clerks, for instance, intelligence tests and number and name comparison

tests are about equal in validity. However, if either of these tests is used alone, it is more than likely that it does not select the same sort of clerical worker as does the other test. It is quite possible that with proper controls of these factors the clerical test might be superior in selecting for lower-level positions, and that as the responsibility of the clerical position becomes greater, the value of an intelligence measure would likewise increase. There is, however, little available information from civilian research which clearly demonstrates this to be so, although it is possible that many selection procedures are based on an implicit acceptance of such an hypothesis. However, none of the studies reported here explicitly differentiated between the levels of clerical job being studied nor attempted to trace any variation in the relationship between success on the job and scores on the tests through the various levels of clerical work. It is impossible to conduct any further analysis of the reported results as no descriptions of the complexity or responsibility of the work were provided.

B. Industrial Workers

Typically studies in this area have involved males as subjects. In the few instances where women have been studied, the jobs have tended to involve tasks such as sewing machine operation or inspection-packing, for which there are no Navy counterparts (Ghiselli (70), Maher and Fife (124), Blum and Candee (31, 32), and Grauer (79)). To the extent that published research adequately reflects the type of employment found by women in industry, one is led to the conclusion that industrial jobs tend to be held by one sex or the other with little intermingling of the sexes on one job within one industry or plant.

As a result there is almost no reported civilian research which provides a basis for comparisons of the sexes in selection for the same job by the same measure. When the search is limited to selection for jobs with tests similar to those used in the Navy the field is narrowed even more drastically. One small study by Forlano and Kirkpatrick (64) of 20 female radio tube mounters reports the use of the Otis SA along

with two measures of personality. A study of 33 women trainee telephone mechanics reported by Oxlade (140) refers to the use of the Otis, the Progressive Matrices, and the ACER Mechanical Comprehension tests.

Since the criterion in the latter study was a highly theoretical examination, and the size of the sample in each of the studies was small, it is, therefore, difficult to appraise the significance of the results.

A study by Bolanovich (33) presented an analysis of data from the records of 86 Radio Corporation of America Engineering Cadettes (female) who attended a ten months' electronics engineering course at Purdue University preparatory to entering jobs as engineering aides in six manufacturing plants of the company. Grade-point averages showed significant correlations with the Cooperative General Mathematics Test for high school students ($r = .55$), the Wonderlic Personnel Test ($r = .50$), previous school grades ($r = .50$), a fitness rating ($r = .38$), and a personality rating ($r = .32$). A maximum shrunken multiple correlation of .61 was found between grade-point averages and a combination of score on the ACE mathematics test and previous school grades. Other tests failed to raise this correlation.

VI. Civilian Studies of the Selection of Women for Navy-like Jobs Using Tests Unlike Those in the Navy Basic Test Battery

A. Clerical Workers

Tests of typing ability are frequently used in selecting clerical workers. Giese (74) reported a correlation of .64 between scores on the Kimberley-Clark Typing Ability Analysis Test and ratings by supervisors for 24 clerical employees. Barrett (11) found that total scores from the Turse Shorthand Aptitude Test differentiated between "good" and "poor" typing students (N = 96) as measured by grades; transcription and phonetic association subscores from the same test were stated to differentiate between "good" and "average" stenography students (N = 75). Holmes (96) reported validity coefficients for tests of typing and shorthand against the criterion of pooled supervisors' ratings. These were .36 and .28 respectively for 88 secretaries; .35 and .29 for 56 typists; and .14 and -.18 for 107 clerk typists.

Several studies have attempted to use personality measures as predictors of success in clerical work, e.g., Dodge (56, 57) and McMurry (123), but are not considered relevant to this discussion and hence will not be discussed in detail.

B. Industrial Workers

The selection of women for the industrial jobs which they most commonly hold, such as inspecting, packing, assembling, and operating power sewing machines, tends to involve measures of special ability quite unlike the tests used by the Navy. Many manual dexterity tests, such as the Minnesota Placing and Turning Test and the tests designed by MacQuarrie and O'Connor, have been used with a moderate degree of success. Reported validity coefficients range from .34 to .62 and appear to cluster around .54. Cross validation, however, is on the whole conspicuous by its absence.

The studies of power sewing machine operators by Treat (179), Otis (138), Ruch (153) and Glanz (75) will be mentioned only in passing as

there is no comparable Navy job. Similar brief mention is made of the several studies by Walker (189), Ayers (8), Coleman (47) and Kerr (102) on visual factors in job success, and of the work of Bolanovich (34) on the influence of interest testing in reducing factory turnover.

Tiffin and Greenly (175) found that, in two of three groups studied, the O'Connor Finger Dexterity Test picked radio and electric fixture assemblers above average in amount and quality of production on the assembly line. Reported correlations with pooled ratings of efficiency were .33 for a group of 33 and .20 for a group of 42 operators. A hand precision test picked operators who were above average, particularly in quality of work, yielding an "r" of .63 for one group of 36 operators and of .24 for another group of 33 operators. For a third group of 42 radio assemblers the same test was useful ($r = .23$) when error score alone was used. Performance on vision tests varied inversely with production. A best-weighted combination of scores from the finger dexterity test, hand precision test, visual acuity, and color vision tests was reported to yield a multiple correlation with pooled ratings of efficiency of .60 for the third group of 42 radio assemblers.

Tiffin and Rogers (177) reported findings for 150 inspectors engaged in examining the quality of tin plate. For this sample, visual discrimination, height, and weight were as important as manual dexterity. Other authors have reported greater success with manual dexterity tests. Rusmore (154) reported validity coefficients of .49 and .60 with ratings of supervisors in jobs of inspecting, labelling, and packaging based on a sample of 28 women. Two studies of the relationship between scores on the MacQuarrie test and ratings of 329 radio assemblers were conducted by Goodman (76, 77). In one study (77) validities were not reported but were considered by the author to be significant. In the other study (76) Goodman reported a zero-order correlation of .42 between instructor's ratings and total scores on the test and a multiple correlation of .46 with scores on the subtests of the MacQuarrie.

In Surgent's (166) study of 233 radio-tube mounters in which the criterion was the pooled ratings of the supervisor and the instructors,

validity coefficients were .56 for the Minnesota Placing Test, .50 for the Minnesota Turning Test, .48 for the O'Connor Tweezer Dexterity Test, and .64 for the Purdue Pegboard Assembly Test. Optimal combination of the tests produced a multiple R of .76.

Blum (30) reported a significant relationship between scores on the O'Connor Finger and Tweezer Dexterity tests and success on the job for applicants accepted in a watch factory. As a result of a similar study, Candee and Blum (44) suggested the use of the Tweezer Dexterity Test in the initial selection of workers for this factory and the use of the Finger Dexterity Test in selecting superior workers. Similarly, Hines and O'Connor (93) reported the successful use of the O'Connor Finger Dexterity Test in selecting women for jobs involving fine meter or instrument work in the West Lynn plant of the General Electric Company.

Hayes (92) reported significant differences in scores on two pegboards (one of which was the O'Connor pegboard) between workers identified as "quick" and "slow" learners. His findings were based on the study of 1541 women engaged as coil winders, drill and punch press operators, operators of insulating machines, and bench hands. The conclusions, however, need qualifying; groups were dichotomized differently for each of his eight samples, and critical ratios were employed when small sample techniques should have been used in view of the size of the sample.

It is evident that tests of manual dexterity have been used with considerable success in selecting women for a number of industrial positions. There is no evidence available to indicate whether or not such tests would require different interpretation in predicting the success of men on the same jobs. Furthermore, there is no indication in reported results that these tests would be an important addition to the tests currently used in the Navy. Further study would be necessary in order to determine the usefulness of dexterity tests as a supplement to the basic battery in selection for jobs in the Navy.

VII. Sex Differences in Scores on Tests
Like Those in the Navy Basic Test Battery

Norms based on samples of both men and women are available for several tests which are similar to the Navy's tests. It is important to know that significant sex differences have been noted for various tests when considering them for use in predicting future job performance. Normative data alone can provide no solutions to problems of the adequacy of prediction but may, at least, point to areas in which further study is needed. Whenever significant sex differences are observed, questions such as the following immediately become important: Is the test providing a measure of the same skills for women that it does for men? Have the scores the same significance for women that they have for men -- will a woman with a given score on the test perform at the same level on the job as will a man with the same test score?

Some data on sex differences in performance on clerical tests and on tests of mechanical comprehension have therefore been considered important because they may serve to raise these and other questions. Reports citing such findings are included below. It is not, of course, to be implied that absence of any report of sex differences on a test indicates that the test is successfully measuring the same ability or skill to the same extent and in the same way for both men and women.

A. Clerical Tests

Of the many tests which purport to measure clerical aptitude, few are provided with separate norms for men and women. Separate normative data based on adequate samples of both sexes are available for only two tests, the Minnesota Clerical Test and the Psychological Corporation's General Clerical Test.

The norms which accompanied the publication in 1933 of the Minnesota Clerical Test (4) showed that for samples drawn from the general population women were decidedly superior to men on this test. The superiority persisted in studies of clerical workers although it was less marked. In the manual accompanying the 1946 revision (5) the authors

reported that studies carried out in the interim period had confirmed the original findings. Only 16 per cent of men reached or exceeded the median for women in samples presumed to represent the general population, and only 21 per cent of the men reached or exceeded the median for women in samples from clerical populations. The authors noted, however, that these differences tended to disappear when comparisons were limited to a given type or level of clerical job. The data presented in the manual were based in part upon research carried out by Loevinger (116), Thatcher (171), and Schneidler and Paterson (155). These studies traced the sex differences back to the fifth grade level. Engelhardt (62) has noted similar differences for college groups.

A similar pattern of sex differences is observed in the published norms for the Psychological Corporation's Clerical Test (144). Women are superior on the average to men on the verbal and clerical subscores and on the total score; the differences, however, are not large. Men tend to score higher on the numerical subtests (for which there is no counterpart in the Minnesota Clerical Test). The norms for the Australian Council for Educational Research Speed and Accuracy Test (6) show women to be superior at every age level for which data was available.

A review of manual for other clerical tests reveals that norms are frequently based entirely upon the testing of female samples. This is true for the Bennett Stenographic Aptitude Test (15) and the Short Employment Tests (20). Other test manuals present one set of norms for both sexes with no reference to the sex of the original sample. This is the case for the ERC Stenographic Aptitude Test (55), the Acorn Clerical Aptitude Test (104), the Chicago Test of Clerical Promise (132), and the NIIP Clerical Test: American Revision (131).

Few of the authors discussed the possible origins of the reported sex differences. Hypotheses have been offered with extreme caution; those, like Schneidler and Paterson (155), who have ventured to speculate at all have demonstrated an interest in the possible influences of differences in training and in encouragement to achieve in these areas.

B. Tests of Mechanical Comprehension

The most striking reports of sex differences were those for tests which were designed to measure mechanical aptitude. Bennett and Cruikshank (17) in a study of 390 girls and 338 boys of comparable age and education found that the boys scored significantly higher on the Test of Mechanical Comprehension through grades ten, eleven, twelve, and into the first year of college. As a result of an item analysis these authors found that certain items discriminated between the sexes to a far greater extent than others. This discovery probably led to the publication in 1951 of form W1 of the Bennett test (19), consisting of items which supposedly do not discriminate against women.

In publishing norms for a test somewhat similar to the Bennett, the Australian Council for Educational Research (7) reported similar findings. In samples of 2,000 adult males and 1,000 adult females chosen to represent the general population, and for university samples of 420 males and 295 females, males scored higher than females. This sex difference was not found in studies using the Detroit Mechanical Aptitudes Examination (9). The last test, however, contains arithmetic, motor ability, and assembly subtests, as well as a subtest of mechanical knowledge; it is, therefore, hardly comparable with the Bennett type of test.

Frederiksen (67) reported a difference in mean score of 3.95 points between a sample of 4857 men and 1340 women on the Mechanical Comprehension Test from the U. S. Navy's Officer Qualification Test. The men displayed a slightly greater variability in score with a standard deviation of 3.63 as opposed to 2.93 for the women.

McElheny (120) related scores on Form AA of the Bennett test to scores on the Purdue Mechanical Assembly Test and also gave some information on the relative performance of the two sexes. Eighty male and twenty female college students were studied. The mean score for males indicated better performance than did that for females on both the Bennett and the Purdue tests.

The Minnesota Paper Form Board Test was stated by Likert and Quasha (115) to be the only test in the Minnesota Mechanical Ability Battery which gave satisfactory correlations with a criterion of mechanical ability. Sex differences are slight, although men tend to excel women according to Alteneder (3) and Tuckman (181). In a study of 1008 Pratt Art School freshmen, Bryan (40) found the mean to be the same (44.5) for both men and women, though the standard deviation was smaller (9.4) for the men than for the women (12.4).

In a study by Stephens (165) based on mid-year measurements of 1797 female and 1139 male seniors in New England high schools, the males scored consistently higher than the females on the Minnesota Paper Form Board Test, although the difference was very small. In a study of 25 men and 25 women college students Bates, Wallace, and Henderson (13) found no large sex differences in performance on this test.

C. Tests of Intelligence

It is not the purpose of this report to deal with the extensive literature that has accumulated in the last quarter of a century on the question of sex differences in intelligence. Attitudes have changed, opinions are today less dogmatically expressed, and as Kuznets and McNemar (106) have stated, when large unselected groups are used, age is taken into account, and possibilities of bias in the test content are allowed for, startling differences in average tendency or in variation simply fail to emerge.

VIII. Implications

As data have accumulated over the past forty years, earlier conceptions of the nature of sex differences in ability have slowly changed or been abandoned. As compared with earlier writings on the topic, there is a noticeable tendency in the recent literature to attribute measured sex differences more to differences in cultural influences than to differences in inherited endowment. In the measurement of intelligence this newer viewpoint has had for some time a considerable impact on test development. The hypothesis that women may be considered as subject to different cultural pressures than men has, however, only recently influenced approaches to the measurement of special abilities such as mechanical aptitude.

This lack of influence may be attributed in part to the rather consistent patterns of employment that have in the past prevailed for men and women. As women in great numbers have not ordinarily sought jobs which demand considerable mechanical aptitude, the presence of sex differences on measures of this aptitude has tended to be of little concern to the practical psychologist. Even in times of emergency women have been channeled into tasks in which their traditionally-conceived superiority in manual dexterity or attention to clerical detail might be advantageously used, as in simple assembly work, inspection, power sewing-machine operation, or clerical work in the armed forces.

In the last two decades women have become a significant part of the labor force in an increasingly wider range of occupations. However, the frequency of their employment, rather than the nature of it, seems to have been the more striking change. The poorer performance of women on tests of arithmetic reasoning and mechanical aptitude has merely served to reinforce a well-conceived and time-honored view of the appropriateness of using sex as one basis for dividing the labor of the country. The arbitrariness of such a view, with its attendant assumption of sex difference in endowment, was only called in question with the advent of an unusually disastrous war, and even then entirely as a result of hindsight.

The British, for instance, selected women to operate anti-aircraft guns and equipment and to serve as tinsmiths and pipe-fitters only because there were no men available to do these jobs. The follow-up data gathered and analyzed indicated that these and other jobs hitherto performed only by men had been performed quite creditably by women.

The British used women in real and stressful work situations but reported little data that would permit a comparative analysis of the relation of achievement to selection test scores for men and women. This lack of information is partly due to the fact that the British wisely side-stepped the problem by making no attempt to select women with exactly the same tests and minimum cutting scores that were used for men.

There is some indication from the British studies that, properly chosen, tests may be used with the same success with women as with men in the selection of personnel for jobs demanding mechanical skills. The studies reported by Wickham (191) and by Vernon and Parry (186) indicated that an assembly test was a useful supplement to the Bennett Test of Mechanical Comprehension in selecting women for such jobs. Similarly, there was some indication from the British research that on-the-job validities of the tests corresponded closely with validities from the training-school situations.

To date, the U. S. Air Force seems to have made the greatest strides in tracing the relationship between scores on selection tests, administered to both men and women, and performance in technical schools. In studying situations in which the same tests were given to members of both sexes who then were trained in the same setting, it appeared that women tended to obtain higher training grades than did men having the same scores on the aptitude measures. This finding may be interpreted in several ways, as Gordon (78) has suggested. Ignoring the possible sources of bias in the criterion measures, it may be said that women were under-measured by the selection tests, or that they over-achieved. It may also be possible that both hypotheses are true in part. In any case, it is evident that a score on the Airman Classification Battery has a different meaning for a woman than it has for a man.

It is reasonable to argue that most selection tests for technical schools in the Air Force tend to measure information about, and familiarity with, particular technical operations. In our culture women are less likely to be acquainted with such operations than are men. However, since the selection variables do show positive correlations with the criterion for either sex taken alone, we must suppose that measures of information have, in fact, some relationship to aptitude, in that those who are talented tend to show interest and hence gain information. The problem would seem to be that of devising a selection procedure which would tap aptitude in women and men without placing at a disadvantage those whose opportunities for familiarity with certain materials or subject matter have been limited.

It might be possible to achieve this end result in three different ways. In the first place, different tests might be used with women; this method was employed by the British with reasonable success in several cases. Second, a study might be made of the regression of performance on selection test scores for the two sexes, as the Air Force has attempted to do. This might lead to use of different minimum qualifying scores for women than for men. Third, it might be possible to devise new tests which measure each particular aptitude in the same way for both sexes; for instance, the speed of acquiring a new skill in a standardized learning situation might be a measure that is nearly independent of cultural influences.

The British and U. S. Air Force findings cited in this report clearly indicate the need for further study of the adequacy of present tests in selecting women for jobs traditionally performed by men. The available evidence, while far from being definitive, still seems sufficient to call into question the use of identical tests and test procedures with women and with men. With present tests it would appear desirable to take sex into account when considering the meaning of a given test score during the process of assigning personnel. The available data on test-criterion relationships seem better described by the diagrams presented as Figures 2 and 3 than by Figure 1. (See pages 4 and 5.)

One further point seems eminently clear: there is a definite need for further follow-up studies of men and women who (1) take the same tests, (2) work on the same jobs, and (3) are evaluated in the same ways.

BIBLIOGRAPHY

Note: Articles referred to in the body of the report are preceded by an asterisk.

1. Abt, L. E. A test battery for selecting technical magazine editors. Personnel Psychol., 1949, 2, 75-91.
2. Allen, C. N. Studies in sex differences. Psychol. Bull., 1927, 24, 294-304.
- *3. Altener, Louise E. The value of intelligence, personality, and vocational interest tests in a guidance program. J. educ. Psychol., 1940, 31, 449-459.
- *4. Andrew, Dorothy M., Paterson, D. G., and Longstaff, H. P. Minnesota Clerical Test. New York: The Psychological Corporation, 1933.
- *5. Andrew, Dorothy M., and Paterson, D. G. Minnesota Clerical Test. New York: The Psychological Corporation, 1946.
- *6. Australian Council for Educational Research. A.C.E.R. Speed and Accuracy Tests. Melbourne: Australian Council for Educational Research, 1942.
- *7. Australian Council for Educational Research. A.C.E.R. Mechanical Comprehension Test. Melbourne: Australian Council for Educational Research, 1942.
- *8. Ayers, A. W. A comparison of certain visual factors with the efficiency of textile inspectors. J. appl. Psychol., 1942, 26, 812-827.
- *9. Baker, H. J., Voelker, P. H., and Crockett, A. C. Detroit Mechanical Aptitudes Examination, Revised, Form A. Bloomington, Ill.: Public School Publishing Co., 1928-39.
10. Baldwin, E. F., and Smith, L. F. The performance of adult female applicants for factory work on the Likert-Quasha revision of the Minnesota Paper Form Board Test. J. appl. Psychol., 1944, 28, 468-470.
- *11. Barrett, Dorothy M. Prediction of achievement in typewriting and stenography in a liberal arts college. J. appl. Psychol., 1946, 30, 624-630.
12. Bartelme, Phyllis F., Fletcher, E. D., Brown, C. W., and Ghiselli, E. E. The prediction of driving skill. J. appl. Psychol., 1951, 35, 98-100.
- *13. Bates, J., Wallace, M., and Henderson, M. T. A statistical study

- of four mechanical tests. Proc. Iowa Acad. Sci., 1943, 50, 299-301. (As abstracted by B. Wellman in Psychol. Abstr., 1944, 18, 3276.)
- *14. Bellows, R. M. Studies of clerical workers. In W. H. Stead, C. L. Shartle, et al., Occupational counseling techniques. New York: American Book Co., 1940, pp. 137-154.
- *15. Bennett, G. K. Stenographic Aptitude Test. New York: The Psychological Corporation, 1937.
16. Bennett, G. K. Distribution of scores from revisions of Army Alpha. J. appl. Psychol., 1943, 27, 150-151.
- *17. Bennett, G. K., and Cruikshank, Ruth M. Sex differences in the understanding of mechanical problems. J. appl. Psychol., 1942, 26, 121-127.
18. Bennett, G. K., and Fear, R. A. Mechanical comprehension and dexterity. Personnel J., 1943-44, 22, 12-17.
- *19. Bennett, G. K., and Fry, Dinah E. Test of Mechanical Comprehension Form W-1. New York: The Psychological Corporation, 1942.
- *20. Bennett, G. K., and Gelink, Marjorie. Short Employment Tests. New York: The Psychological Corporation, 1951.
21. Bennett, G. K., Seashore, H. G., and Wesman, A. G. A manual for the Differential Aptitude Tests. (2nd Ed.) New York: The Psychological Corporation, 1952.
22. Berg, I. A. A study of success and failure among student nurses. J. appl. Psychol., 1947, 31, 389-396.
23. Betts, G. L. (Ed.) Minnesota Rate of Manipulation Test, revised manual. Minneapolis: Educational Test Bureau, 1946.
- *24. Bills, Marian A. Methods for the selection of comptometer operators and stenographers. J. appl. Psychol., 1921, 5, 275-283.
- *25. Bills, Marian A. A test for use in selection of stenographers. J. appl. Psychol., 1921, 5, 373-377.
26. Bills, Marian A. Relation of mental alertness test score to positions and permanency in company. J. appl. Psychol., 1923, 7, 154-156.
27. Bird, Norma. Relationships between experience factors, test scores and efficiency. Arch. Psychol., 1931, 19, No. 126.

- *28. Blakemore, Arline. Reducing typing costs with aptitude tests.
Personnel J., 1951-52, 30 (1), 20-24.
- 29. Blankenship, A. B., and Taylor, H. R. Prediction of vocational proficiency in three machine operations. J. appl. Psychol., 1938, 22, 518-526.
- *30. Blum, M. L. A contribution to manual aptitude measurement in industry. The value of certain dexterity measures for the selection of workers in a watch factory. J. appl. Psychol., 1940, 24, 381-416.
- *31. Blum, M. L., and Candee, Beatrice. The selection of department store packers and wrappers with the aid of certain psychological tests. J. appl. Psychol., 1941, 25, 76-85.
- *32. Blum, M. L., and Candee, Beatrice. The selection of department store packers and wrappers with the aid of certain psychological tests. II. J. appl. Psychol., 1941, 25, 291-299.
- *33. Bolanovich, D. J. Selection of female engineering trainees. J. educ. Psychol., 1944, 35, 545-553.
- *34. Bolanovich, D. J. Interest tests reduce factory turnover. Personnel Psychol., 1948, 1, 81-92.
- 35. Book, W. F., and Meadows, J. L. Sex differences in 5925 high school seniors in ten psychological tests. J. appl. Psychol., 1928, 12, 56-81.
- *36. Brown, C. W., and Ghiselli, E. E. The relationship between the predictive power of aptitude tests for trainability and for job proficiency. J. appl. Psychol., 1952, 36, 370-372.
- 37. Brown, C. W., and Ghiselli, E. E. Prediction of labor turnover by aptitude tests. J. appl. Psychol., 1953, 37, 9-12.
- 38. Brown, C. W., and Ghiselli, E. E. Some generalizations concerning the validity of aptitude tests. Personnel Psychol., 1953, 6, 139-150.
- 39. Brush, E. N. Mechanical ability as a factor in engineering aptitude. J. appl. Psychol., 1941, 25, 300-312.
- *40. Bryan, Alice I. Grades, intelligence, and personality of art school freshmen. J. educ. Psychol., 1942, 33, 50-64.

41. Burr, Emily. Psychological tests applied to factory workers. Arch. Psychol., 1922, 8, No. 55.
42. Burt, H. E. Professor Munsterberg's vocational tests. J. appl. Psychol., 1917, 1, 201-213.
43. California Test Bureau. MacQuarrie Test for Mechanical Ability: Summary of investigations, No. 2. Los Angeles: California Test Bureau, 1950.
- *44. Candee, Beatrice, and Blum, M. L. Report of a study done in a watch factory. J. appl. Psychol., 1937, 21, 572-582.
45. Case, H. W. Selection of aircraft engineering draftsmen and designers. J. appl. Psychol., 1947, 31, 583-588.
46. Clarke, A. C., and Cook, P. H. A group study of women workers. Bull. industr. Psychol. Personnel Pract., Melbourne, 1949, 5 (1), 29-38.
- *47. Coleman, J. H. The visual skills of precision instrument assemblers. J. consult. Psychol., 1945, 9, 165-170.
48. Cook, D. W. Psychological tests for unskilled jobs. Personnel Series No. 50. New York: American Management Association, 1941, 18-29.
49. Cook, D. W., and Barre, Marguerite F. The effect of specialized industrial norms on the use of the Minnesota Rate of Manipulation Test as a selective instrument in employment procedure. J. appl. Psychol., 1942, 26, 785-792.
50. Copeland, H. A. Validating two tests for census enumeration. J. appl. Psychol., 1937, 21, 230-232.
51. Corris, R. C. Testing before training. Personnel J., 1936, 15, 87-90.
52. Crawford, J. E. A test for tridimensional structural visualization: A new test for mechanical insight designed primarily to measure ability or aptitude in drafting. J. appl. Psychol., 1940, 24, 482-492.
53. Davidson, C. M. Analysis of clerical tests. Personnel J., 1937, 16, 95-98.
54. Davidson, C. M. Evaluation of clerical tests. Personnel J., 1937, 16, 57-64.

- *55. Deemer, W. L., Jr. E.R.C. Stenographic Aptitude Test. Chicago: Science Research Associates, 1938, 1944.
- *56. Dodge, A. F. What are the personality traits of the successful clerical worker? J. appl. Psychol., 1940, 24, 576-586.
- *57. Dodge, A. F. Characteristics of good clerks. Personnel J., 1941-42, 20 (9), 324-327.
- 58. Doppelt, J. E., and Bennett, G. K. Reducing the cost of training satisfactory workers by using tests. Personnel Psychol., 1953, 6, 1-8.
- 59. Drake, C. A. Spot the best by test. Factory Mgmt. Maint., 1937, 95 (August), 60-62.
- 60. Dulsky, S. G., and Krout, M. H. Predicting promotion potential on the basis of psychological tests. Personnel Psychol., 1950, 3, 345-351.
- 61. Edgerton, H. A., and Britt, S. H. Sex differences in the science talent test. Science, 1944, 100, 192-193.
- *62. Engelhardt, Olga E. de Cillis. The Minnesota Clerical Test: Sex differences and norms for college groups. J. appl. Psychol., 1950, 34, 412-414.
- 63. Ford, A. H. Prediction of academic success in three schools of nursing. J. appl. Psychol., 1950, 34, 186-189.
- *64. Forlano, G., and Kirkpatrick, F. H. Intelligence and adjustment measurements in the selection of radio tube mounters. J. appl. Psychol., 1945, 29, 257-261.
- 65. Foster, W. R. Vocational selection in a chocolate factory. J. natl. Inst. Industr. Psychol., 1924-25, 2, 159-163.
- 66. Frasier, G. W. A comparative study of the variability of boys and girls. J. appl. Psychol., 1919, 3, 151-155.
- *67. Frederiksen, N. Preparation of norms for the Officer Qualification Test, Form I. (OSRD, 1943: Publ. Bd., No. 13310.) Washington, D. C.: U. S. Dep. Commerce, 1946.
- 68. Fruchter, B. Ability patterns in technical training criteria. J. appl. Psychol., 1952, 36, 381-384.
- 69. Gadel, Marguerite S., and Kriedt, P. H. Relationships of aptitude, interest, performance, and job satisfaction of IBM operators. Personnel Psychol., 1952, 5, 207-212.

- *70. Ghiselli, E. E. Tests for the selection of inspector-packers.
J. appl. Psychol., 1942, 26, 468-476.
- 71. Ghiselli, E. E., and Barthol, R. P. The validity of personality inventories in the selection of employees. J. appl. Psychol., 1953, 37, 18-20.
- *72. Ghiselli, E. E., and Brown, C. W. Validity of aptitude tests for predicting trainability of workers. Personnel Psychol., 1951, 4, 243-260.
- 73. Ghiselli, E. E., and Brown, C. W. Validity of tests for auto mechanics. J. appl. Psychol., 1951, 35, 23-24.
- *74. Giese, W. J. A tested method for the selection of office personnel. Personnel Psychol., 1949, 2, 525-545.
- *75. Glanz, E. A trade test for power sewing machine operators. J. appl. Psychol., 1949, 33, 436-441.
- *76. Goodman, C. H. The MacQuarrie Test for Mechanical Ability: I. Selecting radio assembly operators. J. appl. Psychol., 1946, 30, 586-595.
- *77. Goodman, C. H. The MacQuarrie Test for Mechanical Ability: III. Follow-up study. J. appl. Psychol., 1947, 31, 502-510.
- *78. Gordon, Mary A. A study in the applicability of the same minimum qualifying scores for technical schools to white males, WAF, and negro males. Human Resources Research Center, Air Research and Development Command, Lackland Air Force Base, Texas, Technical Report 53-34, November, 1953.
- *79. Grauer, D. Abilities, attitudes, and success: A study of sewing-machine operators. Personnel J., 1933-34, 12 (6), 328-333.
- 80. Gulliksen, H. Preparation of norms for women for the Officer Qualification Test, Form 1. (OSRD, 1943; Publ. Bd., No. 13308.) Washington, D. C.: U. S. Dep. Commerce, 1946, pp. 3.
- 81. Habbe, S. The selection of student nurses. J. appl. Psychol., 1933, 17, 564-580.
- 82. Halliday, R. W., Fletcher, F. M., Jr., and Cohen, Rita M. Validity of the Owens-Bennett Mechanical Comprehension Test. J. appl. Psychol., 1951, 35, 321-324.

83. Hanman, B. The performance of adult males on the Minnesota Paper Form Board Test and the O'Rourke Mechanical Aptitude Test. J. appl. Psychol., 1942, 26, 809-811.
84. Harrell, W. A factor analysis of mechanical ability tests. Psychometrika, 1940, 5, 17-33.
85. Harrell, W. Testing cotton mill supervisors. J. appl. Psychol., 1940, 24, 31-35.
86. Harris, A. J. The relative significance of measures of mechanical aptitude, intelligence, and previous scholarship for predicting achievement in dental school. J. appl. Psychol., 1937, 21, 513-521.
87. Harrison, R., and Jackson, T. A. Validation of a clinical approach to the placement of engineers. J. appl. Psychol., 1952, 36, 373-376.
- *88. Hay, E. N. Predicting success in machine bookkeeping. J. appl. Psychol., 1943, 27, 483-493.
- *89. Hay, E. N. Cross-validation of clerical aptitude tests. J. appl. Psychol., 1950, 34, 153-158.
- *90. Hay, E. N. Mental ability tests in clerical selection. J. appl. Psychol., 1951, 35, 250-251.
91. Hay, E. N. Some research findings with the Wonderlic Personnel Test. J. appl. Psychol., 1952, 36, 344-345.
- *92. Hayes, Elinor G. Selecting women for shop work. Personnel J., 1932-33, 11, 69-85.
- *93. Hines, Mildred, and O'Connor, J. A measure of finger dexterity. J. Personnel Res., 1925-26, 4, 379-382.
94. Hobson, J. R. Sex differences in primary mental abilities. J. educ. Res., 1947, 41, 126-132.
- *95. Holmes, F. J. Validity of tests for insurance office personnel. Personnel Psychol., 1950, 3, 57-69.
- *96. Holmes, F. J. Validity of tests for insurance office personnel, II. Personnel Psychol., 1950, 3, 217-220.
- *97. Howard, J. K., and Pickrel, E. W. Validation of the Airman Classification Battery for women in the Air Force. Human Resources Research Center, Air Training Command, Lackland Air Force Base, San Antonio, Texas. Technical Report 52-5, November, 1952.

98. Jaspen, N. A factor study of worker characteristics. J. appl. Psychol., 1949, 33, 449-459.
99. Jurgensen, C. E. Extension of the Minnesota Rate of Manipulation Test. J. appl. Psychol., 1943, 27, 164-169.
100. Jurgensen, C. E. Norms for the Test of Mechanical Comprehension. J. appl. Psychol., 1948, 32, 618-621.
101. Kephart, N. C. Visual skills and labor turnover. J. appl. Psychol., 1948, 32, 51-55.
- *102. Kerr, W. A. Vision tests for precision workers at R.C.A. Personnel Psychol., 1948, 1, 63-66.
- *103. Kinney, L. B. The relationship of certain factors to success in clerical positions. J. appl. Psychol., 1933, 17, 55-62.
- *104. Kobal, A., Wrightstone, J. W., Kunze, K. R., and MacElroy, A. J. Acorn Clerical Aptitude Test. New York: Acorn Publishing Company, 1943.
105. Kriedt, F. H., and Gadel, Marguerite S. Prediction of turnover among clerical workers. J. appl. Psychol., 1953, 37, 338-340.
- *106. Kuznets, G. M., and McNemar, Olga. Sex differences in intelligence-test scores. 39th Yearbook, National Soc. Stud. Educ., 1940. Pt. I, 211-220.
107. Laney, A. R., Jr. Validity of employment tests for gas-appliance service personnel. Personnel Psychol., 1951, 4, 199-208.
108. Lawshe, C. H., Jr. The Purdue Industrial Training Classification Test. J. appl. Psychol., 1942, 26, 770-776.
109. Lawshe, C. H., Jr. How can we pick better supervisors? Personnel Psychol., 1949, 2, 69-73.
110. Lawshe, C. H., Jr., and McGinley, A. D., Jr. Job performance criteria studies: I. The job performance of proofreaders. J. appl. Psychol., 1951, 35, 316-320.
111. Lawshe, C. H., Jr., Semanek, Irene A., and Tiffin, J. The Purdue Mechanical Adaptability Test. J. appl. Psychol., 1946, 30, 442-453.
112. Lawshe, C. H., Jr., and Thornton, G. R. A test battery for identifying potentially successful Naval electrical trainees. J. appl. Psychol., 1943, 27, 399-406.

113. Lawshe, C. H., Jr., and Tiffin, J. How tests can strengthen the training program. Factory Mgmt Maint., 1944, 102 (March), 119-121.
114. Lee, Marilyn C. Relationship of masculinity-femininity to tests of mechanical and clerical abilities. J. appl. Psychol., 1952, 36, 377-380.
- *115. Likert, R., and Quasha, W. H. Revised Minnesota Paper Form Board Test, Series AA and Series EB: A test of mechanical ability. New York: The Psychological Corporation, 1941, 1948.
- *116. Loevinger, J. An analysis of verbal and numerical abilities at the junior high school level. Unpublished master's thesis, University of Minnesota, 1938. (Cited in Schneider, G. R. and Paterson, D. G. (155) and in Andrew, Dorothy M., and Paterson, D. G. (5).)
117. Louttit, C. M. A bibliography of sex differences in mental traits. Train. Sch., 1925-26, 22, 129-138.
118. McDaniel, J. W., and Reynolds, W. A. A study of the use of mechanical aptitude tests in the selection of trainees for mechanical occupations. Educ. psychol. Measmt., 1944, 4, 191-197.
119. Macdonald, G. L. Measuring progress in radio training. J. appl. Psychol., 1945, 29, 301-312.
- *120. McElheny, W. T. A study of two techniques of measuring "mechanical comprehension." J. appl. Psychol., 1948, 32, 611-617.
121. MacElroy, A. J. (Ed.) Acorn National Aptitude Tests: Clerical Aptitude Test manual of directions. Rockville Centre, New York: Acorn Publishing Company, 1943.
122. McGehee, W., and Moffie, D. J. Psychological tests in the selection of enrollees in engineering, science, management, defense training courses. J. appl. Psychol., 1942, 26, 584-586.
- *123. McMurry, R. N. Efficiency, work-satisfaction and neurotic tendency: A study of bank employees. Personnel J., 1932-33, 11, 201-210.
- *124. Maher, H., and Fife, Isabelle E. A biological-pharmaceutical checker selection program. J. appl. Psychol., 1947, 31, 469-476.

- *125. Marcus, L. Vocational selection for specialized tasks: A study of selective tests for Hollerith-machine operatives. J. appl. Psychol., 1920, 4, 186-201.
- 126. Martin, G. C. Test batteries for trainees in auto mechanics and apparel design. J. appl. Psychol., 1951, 35, 20-22.
- 127. Martin, H. G. Predicting trainability in the use of the Burroughs adding machine. Educ. psychol. Measmt., 1952, 12, 699-700.
- *128. Mercer, Edith O. Psychological methods of personnel selection in a women's service. Occup. Psychol., 1945, 19, 180-200.
- 129. Miller, D. R. Survey of mechanical insight. Los Angeles: California Test Bureau, 1945.
- 130. Moore, B. V. Analysis of results of tests administered to men in the engineering defense training courses. J. appl. Psychol., 1941, 25, 619-635.
- *131. Moore, H. The N.I.I.P. Clerical Test: American revision. New York: The Psychological Corporation (distributors), 1934.
- *132. Moore, H. The Chicago Test of Clerical Promise. Chicago: Science Research Associates, 1941.
- 133. Munro, M. S., and Raphael, Winifred. Selection tests for clerical occupations. J. natl. Inst. industr. Psychol., 1930-31, 5, 127-137.
- *134. Oberheim, Grace M. The prediction of success of student assistants in college library work. Educ. psychol. Measmt., 1942, 2, 379-385.
- *135. Oberheim, Grace M. The relationship between scores on a clerical test and clerical proficiency in library work. J. educ. Psychol., 1944, 35, 493-499.
- 136. O'Rourke, L. J. Office employment tests. Office Management Series No. 46. New York: American Management Association, 1930, 2-21.
- 137. O'Rourke, L. J. O'Rourke Mechanical Aptitude Test, Form A. Washington, D. C.: The Psychological Institute, 1940.
- *138. Otis, J. L. The prediction of success in power sewing machine operating. J. appl. Psychol., 1938, 22, 350-366.

139. Owens, W. A., Jr. A difficult new test of mechanical comprehension. J. appl. Psychol., 1950, 34, 77-81.
- *140. Oxlade, M. An experiment in the use of psychological tests in the selection of women trainee telephone mechanics. Bull. industr. Psychol. personnel Pract., Melbourne, 1946, 2 (1), 26-32.
141. Petrie, Asenath, and Powell, Muriel B. The selection of nurses in England. J. appl. Psychol., 1951, 35, 281-286.
142. Pond, Millicent. Tests for the selection of office employees. Personnel, 1941, 17, 199-208.
143. Porubcn, A., Jr. A test battery for actuarial clerks. J. appl. Psychol., 1950, 34, 159-162.
- *144. The Psychological Corporation. Psychological Corporation General Clerical Test. New York: The Psychological Corporation, 1944.
145. Raven, J. C. The R.E.C.I. series of perceptual tests: An experimental survey. Brit. J. med. Psychol., 1939, 18, 16-34.
146. Remmers, H. H., and Schell, J. W. Testing the O'Connor Wiggly Block Test. Personnel J., 1933, 12, 155-159.
147. Rhinehart, Jesse B. An attempt to predict the success of student nurses by the use of a battery of tests. J. appl. Psychol., 1933, 17, 277-293.
148. Richardson, Bellows, Henry and Co., Inc. SRA Mechanical Aptitudes. Chicago: Science Research Associates, 1947.
- *149. Rogers, H. W. Some empirical tests in vocational selection. Arch. Psychol., 1922, 7, No. 49.
150. Rossett, M. E., and Arakelian, P. A test battery for the selection of panel dial switchmen. J. appl. Psychol., 1939, 23, 358-366.
151. Rossett, M. E., and Arakelian, P. Hiring telephone men by tests. Personnel J., 1939, 18, 184-192.
152. Rothe, H. F. Normative and validity data for the Purdue Mechanical Adaptability Test. Personnel Psychol., 1950, 3, 187-192.
- *153. Ruch, F. L. How to use employment tests. Employment Testing Bulletin No. 1. Los Angeles: California Test Bureau, 1944.
- *154. Ruzmore, J. T. The R-G Pegboard Test of Finger Dexterity. J. appl. Psychol., 1942, 26, 523-529.

- *155. Schneidler, Gwendolen G., and Paterson, D. G. Sex differences in clerical aptitude. J. educ. Psychol., 1942, 33, 303-309.
- 156. Schultz, R. S. Personnel selection in aviation industry. Personnel J., 1940-41, 19, 99-105.
- 157. Seashore, H. B. and Bennett, G. K. A test of stenography: Some preliminary results. Personnel Psychol., 1948, 1, 197-209.
- *158. Seashore, H. G. Validation of clerical testing in banks. Personnel Psychol., 1953, 6, 45-56.
- 159. Seashore, R. H. Individual differences in motor skills. J. gen. Psychol., 1930, 3, 38-66.
- 160. Severin, Daryl. The predictability of various kinds of criteria. Personnel Psychol., 1952, 5, 93-104.
- 161. Shartle, C. L. New selection methods for defense jobs. Personnel Series No. 50. New York: American Management Association, 1941, 30-41.
- 162. Sheddan, B. R., and Witmer, Louise R. Employment tests for relief visitors. J. appl. Psychol., 1939, 23, 270-279.
- 163. Shellow, Sadie M. An intelligence test for stenographers. J. personnel Res., 1926-27, 5, 306-308.
- 164. Stenquist, J. L. Stenquist Mechanical Aptitude Tests. Yonkers-on-Hudson, New York: World Book Company, 1922.
- *165. Stephens, E. W. A comparison of New England norms with national norms on the Revised Minnesota Paper Form Board Test, Series AA. Occupations, 1945, 24, 101-104.
- *166. Surgent, L. V. The use of aptitude tests in the selection of radio tube mounters. Psychol. Monogr., 1947, 61, No. 2 (Whole No. 283).
- 167. Taylor, C. W. Pre-testing saves training costs. Personnel Psychol., 1952, 5, 213-239.
- 168. Taylor, E. K., and Tajen, Claire. Selection for training: Tabulating equipment operators. Personnel Psychol., 1948, 1, 341-348.
- 169. Teegarden, Lorene. Manipulative performance of young adult applicants at a public employment office -- Part I. J. appl. Psychol., 1942, 26, 633-652.

170. Teegarden, Lorene. Manipulative performance of young adult applicants at a public employment office -- Part II. J. appl. Psychol., 1942, 26, 754-769.
- *171. Thatcher, M. Sex differences in clerical ability at fifth and sixth grade levels. Unpublished manuscript, Univer. of Minnesota (Psychol. Dept.), 1940. (Cited in Andrew, Dorothy M., and Peterson, D. G. (5))
172. Thurstone, L. L. A standardized test for office clerks. J. appl. Psychol., 1919, 3, 248-251.
173. Tiffin, J. Industrial Psychology. New York: Prentice-Hall, 1942.
174. Tiffin, J., and Asher, E. J. The Purdue Pegboard: Norms and studies of reliability and validity. J. appl. Psychol., 1948, 32, 234-247.
- *175. Tiffin, J., and Greenly, R. J. Employee selection tests for electrical fixture assemblers and radio assemblers. J. appl. Psychol., 1939, 23, 240-263.
- *176. Tiffin, J., and Lawshe, C. H., Jr. An adaptability test: A fifteen minute mental alertness test for use in personnel allocation. J. appl. Psychol., 1943, 27, 152-163.
- *177. Tiffin, J., and Rogers, H. B. The selection and training of inspectors. Personnel, 1941-42, 18, 14-31.
178. Traxler, A. E. Objective testing in the field of accounting. Educ. psychol. Measmt., 1951, 11, 427-439.
- *179. Treat, Katharine. Tests for garment machine operators. Personnel J., 1929, 8, 19-28.
180. Tuckman, J. Norms for the MacQuarrie Test for Mechanical Ability for high school students. Occupations, 1946, 25, 94-96.
- *181. Tuckman, J. Age and grade norms for high school students on the Revised Minnesota Paper Form Board Test. Occupations, 1947, 26, 97-100.
182. Tuttle, C. W. Women war-workers at Vultee Aircraft. Personnel J., 1942, 21, 82-95.
- *183. Vernon, P. E. Statistical methods in the selection of Navy and Army personnel. Suppl. J. Royal Statist. Soc., 1946, 8, 139-153.

- *184. Vernon, P. E. Psychological tests in the Royal Navy, Army, and A.T.S.. Occup. Psychol., 1947, 21, 53-74.
- *185. Vernon, P. E. Research on personnel selection in the Royal Navy and the British Army. Amer. Psychologist, 1947, 2, 35-51.
- *186. Vernon, P. E., and Parry, J. B. Personnel selection in the British forces. London: University of London Press, 1949.
- 187. Viteles, M. S. What's new in methods for selecting office workers. Office Management Series No. 75. New York: American Management Association, 1936, 4-10.
- 188. Wadsworth, G. W., Jr. Tests prove worth to a utility. Personnel J., 1935, 14, 183-187.
- *189. Walker, W. B. Vision and production of sewing machine operators. Personnel Psychol., 1953, 6, 291-295.
- 190. Weiss, I. Prediction of academic success in dental school. J. appl. Psychol., 1952, 36, 11-14.
- *191. Wickham, Mary. Follow-up of personnel selection in the A.T.S. Occup. Psychol., 1949, 23, 153-169.
- 192. Williamson, E. G., Stover, R. D., and Fiss, C. B. The selection of student nurses. J. appl. Psychol., 1938, 22, 119-131.
- 193. Worbois, G. M. Predicting long-range performance of substation operators. J. appl. Psychol., 1951, 35, 15-19.
- 194. Wrightstone, J. W., and O'Toole, C. E. Prognostic Test of Mechanical Abilities. Los Angeles: California Test Bureau, 1946.

Armed Services Technical Inform

Because of our limited supply, you are requested to return this to
YOUR PURPOSE so that it may be made available to other requesters
will be appreciated.

AD

4626

NOTICE: WHEN GOVERNMENT OR OTHER DRAWINGS, SPECIFICATIONS, OR OTHER DATA ARE USED FOR ANY PURPOSE OTHER THAN IN CONNECTION WITH A GOVERNMENT PROCUREMENT OPERATION, THE U. S. GOVERNMENT ASSUMES NO RESPONSIBILITY, NOR ANY OBLIGATION WHATSOEVER; AND THE GOVERNMENT MAY HAVE FORMULATED, FURNISHED, OR IN ANY WAY SAID DRAWINGS, SPECIFICATIONS, OR OTHER DATA IS NOT TO BE REPRODUCED OR OTHERWISE AS IN ANY MANNER LICENSING THE HOLDER OR PERSON OR CORPORATION, OR CONVEYING ANY RIGHTS OR PERMISSIONS TO REPRODUCE, COPY, OR OTHERWISE TO MANUFACTURE, REPRODUCE, OR OTHERWISE.

REPRODUCED DATA
IN THIS RELATED
DRAWINGS
AND THE
GOVERNMENT
HOLD BY
OR FOR OTHER
TO MANUFACTURE,
REPRODUCE, OR OTHERWISE.

Reproduced by
DOCUMENT SERVICE CENTER
KNOTT BUILDING, DAYTON, 2, OHIO

UNCLASSIFIED