

# PRINCIPLES OF EMPLOYMENT PSYCHOLOGY

BY

HAROLD ERNEST BURTT, PH.D.

*Professor of Psychology, Ohio State University*



HOUGHTON MIFFLIN COMPANY

BOSTON NEW YORK CHICAGO SAN FRANCISCO

*The Riverside Press Cambridge*

COPYRIGHT, 1926  
BY HAROLD ERNEST BURTT  
ALL RIGHTS RESERVED

X : 9 : (S)

F6

5699

The Riverside Press  
CAMBRIDGE · MASSACHUSETTS  
PRINTED IN THE U.S.A.

**Dhananjayrao Gadgil Library**



**GIPE-PUNE-005699**

**TO MY MOTHER**

## PREFACE

THIS book is an outgrowth, in the first instance, of material used for several years in presenting principles of employment psychology to college students, and in the second instance, of practical experience in personnel work and frequent contact with business men interested in psychology in so far as it relates to their problems. Effort is made, on the one hand, to give a fairly comprehensive account of the principles involved for the use of students preparing for practical psychological work in industry, and on the other hand, to avoid a discussion that is too technical for the reader without a psychological background. This does not mean, however, that the treatment is superficial. It is hoped, on the contrary, that the business man reading the book will realize the importance of a careful experimental approach to scientific employment psychology.

Statistical methods must of necessity form a part of the discussion. Although many persons shy at statistics, they are of such wide applicability in employment work that they cannot logically be omitted. No assumption of mathematical knowledge, however, is made, and the effort has been to make any statistical discussion as simple and as clear as possible. Wherever it has proved feasible to describe a method in a general way and relegate the more exacting details to the appendix, this has been done.

The critical psychological reader will notice that no definite stand has been taken regarding the fundamental points of view or metaphysical considerations of theoretical psychology. The author feels that these problems are not germane to the present discussion. The important thing is to predict occupational success whether this is construed from the standpoint of mind or muscle. It is pragmatically justifiable to speak of a "test of attention" regardless of the ultimate nature of attention or whether such a category exists at all. The employment psychologist's task is to arrive at his practical goal regardless of the route taken. Most of us en-



gaged in this field are too busy with our own problems to solve the fundamental issues with which other psychologists are better qualified to deal. In the following discussion it will probably be found that the methods are for the most part objective, but that the terminology is conventional. Experience in presenting psychological principles to the practical man has indicated the desirability of discussing them in terms of everyday vocabulary.

A work of this sort naturally draws rather heavily from experimental material contributed by many psychologists. Such studies as are referred to are cited mainly for illustrative purposes rather than in the nature of a critical review. To this end little mention is made of many details such as the time limit for a mental test, or the actual number of persons involved in a particular experimental study. This is done in order to avoid confusion from too many figures, and does not detract appreciably from the illustrative value of the citation. Reference is always made to the bibliography, however, so that the critical reader may, if he wishes, consult the original article and evaluate it for himself.

The author is indebted to all those who have contributed their bit to the body of psychological knowledge in this general field and whose results have been drawn upon rather extensively. He is especially indebted to H. L. Hollingworth, W. D. Scott, and A. J. Snow, whose contributions have been quoted more extensively. Grateful acknowledgment is likewise made to D. Appleton and Company, New York, for permission to quote from *Vocational Psychology*, by H. L. Hollingworth, *Judging Human Character*, by H. L. Hollingworth, and *Applied Psychology*, by H. L. Hollingworth and A. T. Poffenberger; to the McGraw-Hill Book Company, Inc., New York, for permission to quote from *The Selection and Training of Salesmen*, by H. G. Kenagy and C. S. Yoakum; to James P. Porter, editor, for permission to quote from *The Journal of Applied Psychology*; to A. W. Shaw Company, Chicago, for permission to quote from *Personnel Management*, by W. D. Scott and R. C. Clothier, and *Psychology in Business Relations*, by A. J. Snow; and to The Williams and Wilkins Company, Baltimore, for permission to quote from *Ability to Sell*, by M. J. Ream, and from *The Journal of Personnel Research*.

## PREFACE

vii

It is hoped that the book will show the practical man the importance of painstaking scientific technique in employment psychology in contrast with the expeditious but unreliable methods of unscientific pseudo-psychology. On the other hand, it is hoped that students who expect to pursue psychology in a practical way will find herein a fairly adequate background for plunging further into details.

HAROLD E. BURTT

COLUMBUS, OHIO

# CONTENTS

I. INTRODUCTION . . . . .	1
General psychology — Applied psychology — Psychology in industry — The need for employment psychology — The problem of employment psychology — Fundamental principle of employment psychology — Where employment psychology is most valuable — Employment psychology and human welfare — Outline.	
II. PSYCHOLOGY GOLD BRICKS . . . . .	14
The intellectual underworld — Astrology — Spiritualism — Phrenology — Physiognomy — Experimental evaluation of character analysis from physiognomy — Summary.	
III. HISTORY OF SCIENTIFIC VOCATIONAL PSYCHOLOGY . . . . .	46
Individual differences — Early development of mental tests — Comparison of test scores with occupational ability — Psychology in the war — Employment psychology since the war — Summary.	
IV. TYPES OF MENTAL TESTS . . . . .	60
Classification — Tests of special capacity — Motor control — Sensory capacity — Attention — Learning association — Reaction time — Space perception — Memory — Reasoning — Speed of decision — Ingenuity — Ability to follow directions — Emotion — Honesty — Temperament — Tests of intelligence — Summary.	
V. MENTAL TEST TECHNIQUE . . . . .	101
Method of administration: Individual <i>vs.</i> group tests — Administration: Method of test response — Administration: Types of test response — Administration: Limitation upon test response — Administration: General precautions — Test material — Test instructions — Incentive — Scoring of tests — Reliability of tests — Summary.	
VI. THE CRITERION . . . . .	141
Necessity — Estimates by superiors — Production figures — Miscellaneous criteria — Combining heterogeneous criteria — Summary.	
VII. THE SUBJECTS USED IN EVALUATING TESTS . . . . .	172
General considerations — Applicants for employment as subjects — Employees as subjects — Sampling of subjects — Miscellaneous factors — Summary.	

VIII. SPECIAL CAPACITY TESTS: TOTAL MENTAL SITUATION . . .	185
Preliminary procedure for test research — Analyzing the job with a view to selecting tests — Reproduction of the total mental situation — Giving the tests to workers whose criterion is available — Correlation of test score with criterion — Limitations of total situation method — Examples — Critical scores — Summary.	
IX. SPECIAL CAPACITY TESTS: THE MENTAL COMPONENTS OF THE JOB . . . . .	226
Preliminary selection of tests — Final selection of tests — Weighting the final group of tests — Examples with partial correlation technique — Examples with less statistical refinement — Miscellaneous brief examples — Follow-up procedure — Summary.	
X. INTELLIGENCE AND VOCATIONAL APTITUDE . . . . .	265
Occupational hierarchy — Intelligence surveys — Correlation of intelligence with the criterion — Critical scores in intelligence — Optimum <i>vs.</i> maximum intelligence — Summary.	
XI. INTERESTS IN EMPLOYMENT PSYCHOLOGY . . . . .	294
Permanence of interest — Interest and ability — Methods of measuring interest — Evaluation of measurements of interests — Summary.	
XII. RATING SCALES . . . . .	317
Purpose — Selection of traits to be rated — Weighting the traits — Defining the traits — Man-to-man rating scale — Rating scale for officers — Rating scale for minor executives — Rating by defined groups — Rating scale for executives — Leadership — Graphic rating scale — Reliability of ratings — Validity of ratings — Sources of error in rating procedure — Training rates — Summary.	
XIII. MISCELLANEOUS DETERMINANTS OF VOCATIONAL APTITUDE . . . . .	371
Value — Academic record — Initial and subsequent success in the same occupation — Personal history or application blank — Letters of application — Recommendations — The interview — Interviewer's rating scale for executives — Summary.	
XIV. TRADE TESTS . . . . .	431
Trade tests <i>vs.</i> tests of innate capacity — Need for trade tests — Requirements of trade tests — Principles on which trade tests are based — Kinds of trade tests — Methods of developing and standardizing — Examples of trade tests — General precautions — Summary.	
XV. JOB ANALYSIS . . . . .	467
Nature of job analysis — Purpose — Need — The rôle of psychology in job analysis — Method of securing data — Occupational description — Psychological possibilities in job analysis for employment purposes — Summary.	

## CONTENTS

xi

<b>XVI. THE OUTLOOK FOR EMPLOYMENT PSYCHOLOGY . . . . .</b>	<b>489</b>
Summary of psychological technique applied to employment methods — Present trends — Attitude of workers and management toward employment psychology — Necessity for further research — Essentials for future research — The social implications of employment psychology.	
<b>APPENDIX</b>	
<b>I. ILLUSTRATING THE TECHNIQUE OF CORRELATION . . . . .</b>	<b>511</b>
<b>II. COMBINING RANKINGS WITH INCOMPLETE DATA . . . . .</b>	<b>517</b>
<b>III. ILLUSTRATING THE DERIVATION OF A REGRESSION EQUATION</b>	<b>520</b>
<b>BIBLIOGRAPHY . . . . .</b>	<b>527</b>
<b>INDEX . . . . .</b>	<b>565</b>

# PRINCIPLES OF EMPLOYMENT PSYCHOLOGY



## CHAPTER I INTRODUCTION

**General psychology.** Psychology studies the mind and its objective manifestations. It endeavors to describe facts and to derive general laws with a view to predicting how one will feel, think, and act under given conditions and with a view to controlling those feelings, thoughts, and actions by controlling the conditions. Its method from the time of classical Greece until the nineteenth century was largely *speculative* and casual. The science has now, however, moved from the armchair to the laboratory and is distinctly *experimental* in character. If the early psychologist was interested, for instance, in the bodily accompaniments of emotion, he sat and imagined himself in some dangerous or otherwise emotional situation and tried to observe his bodily feelings. The modern psychologist approaches the same problem by recording on a rotating drum covered with smoked paper the pulse, breathing, blood pressure, and involuntary movements of a person on whom the experiment is being conducted, and then induces emotional states by presentation of pictures, by personal insult, by snakes, or by providing a situation in which the person must sometimes lie and sometimes tell the truth. The early psychologist investigated color vision by looking at the sunset. The modern psychologist throws a beam of light through a prism and with narrow slits selects from the resulting spectrum bands of colored light of known wave length and varies their energy to determine the effect on visibility. To study the process of association of ideas, the early psychologist looked at some object such as a tree and noted what ideas came to him as a result. The modern psychologist uses an instrument

which suddenly exposes a typewritten stimulus word and measures in thousandths of a second the time between the instant of exposure and the instant an observer speaks into a diaphragm the first associated word that comes to him. The early psychologist was content with a few casual observations. The modern psychologist often makes hundreds and subjects the results to rigorous mathematical treatment. In almost any psychological laboratory are to be found precision instruments; adaptations of electrical and mechanical principles to specific problems, printed blanks for standardized mental tests, and statistical equipment.

Applied psychology is a more recent development. Almost no practical use was made of psychological principles until the present century. There are several reasons for this. In the first place there could be no application of the science until there were some principles to apply. A certain theoretical background is necessary in any science before it reaches the practical stage. It must be remembered that, although there was some experimentation prior to that time, the first actual psychological laboratory was established in 1879. There were many psychologists as late as 1917 who believed that the theoretical basis had not been sufficiently laid for an applied science. Most of them, however, aided in the attempt to apply psychology to war problems and their change of opinion was amply justified.

A second factor which delayed the advent of applied psychology was the charlatan. Many a worthless proposition for improving efficiency, analyzing character, or curing ailments was presented under the guise of psychology. People invested in such propositions and were subsequently disappointed. Consequently, when a real applied psychologist approached them they recalled their earlier experience with "psychology" and failed to react favorably to his proposals. The psychology "gold brick" will be discussed in more detail in the next chapter. The point in the present connection is that the pseudo-psychologist injured the reputation of the real psychologist and made it difficult for the latter to make progress in his practical contacts.

A third factor involved in the late development of applied psychology was the emphasis on general laws rather than on individual

differences. Following the lead of the other sciences the general principles were studied first. It was quite natural that interest should first center, for instance, on the general relation between memory and the method by which a poem was studied — as a whole or piecemeal — rather than on the fact that one individual possessed different memory ability from another. It was likewise to be expected that the earlier experiments would be more concerned with determining to which sort of signal one could react more quickly — auditory or visual — rather than with ascertaining whether one individual could react a few hundredths of a second more rapidly than another individual. Yet it is these latter aspects that are often of greatest interest to the applied psychologist. He is concerned with such things as the intelligence of an individual child who is backward in school, the early emotional experiences of a particular patient with an obsession, the changes in blood pressure of a given criminal suspect during examination, or the attention and reaction time of a certain prospective employee. Until there was a partial shift from the study of general laws to the investigation of individual differences the time was not ripe for applied psychology. The last ten or fifteen years, however, have witnessed very distinct advances in the contact of psychology with education, law, medicine, and business.

**Psychology in industry.** Modern industry is especially concerned with three things — raw materials, equipment to construct the product from these materials, and human beings to operate the equipment, keep records, plan, and supervise. The first of these involves such sciences as geology, botany, chemistry, and economics; the second falls especially within the sphere of engineering; but in the third there has developed of late years a realization of the importance of psychology. This importance lies in two outstanding directions: (1) selection of personnel; (2) industrial efficiency. The first of these involves primarily the placement of persons in the type of work to which they are best adapted. The second involves giving the person thus placed a chance to realize his maximum efficiency by proper adjustment of the methods and conditions of work. It involves such problems as training workmen, economy of movement, reduction of fatigue and monotony, the effect upon efficiency



of ventilation or illumination, and industrial harmony. The present book is confined entirely to the first of these two aspects of industrial psychology — the selection of personnel.

The need for employment psychology is obvious. Every employment manager and every foreman is familiar with the occupational misfit — the square peg in the round hole. The explanation of the presence of such misfits in industry is simple. Different jobs require for their satisfactory performance different mental and motor capacities. Individuals differ in mental and motor capacity, and it is frequently the case that the capacity necessary for the job and the capacity possessed by the person working at that job do not correspond. Suppose, to take an oversimplified example, that a job requires good memory and that applicants with good memory and with poor memory are available in about equal numbers. If they are hired at random, about half of them are doomed to failure because they lack the requisite memory ability. A careful survey of almost any large plant would reveal many a workman with slow reaction time vainly trying to keep up with a rapidly operating machine, or a man with poor powers of attention attempting to concentrate on a task that is too complex for him, or with intelligence too low to grasp the problems and make the decisions necessary in his work.

The remedy consists obviously in placing a man in a job requiring aptitudes which he possesses. The management may know pretty well the requirements of the job just as it knows the requirements of the raw materials, but while it measures the tensile strength of the fabric and the specific gravity of the compound, it makes no effort to measure the mind of the workman who is handling that fabric or compound. There was good reason for this a few years ago because methods of mental measurement were not available, but this is no longer true. The development of mental tests, rating scales, and statistical technique has opened up a wide field for scientific contribution to the problems of employment.

The problem of employment psychology consists of determining what mental capacities are needed for a given occupation and devising methods of measuring those capacities. These measurements may then be used upon applicants to determine their prob-

able success in the occupation. Instead of hiring a man without consideration of his mental qualifications and waiting for time to show whether or not he is wisely placed, it is possible at the time of hiring to make some prediction of his ultimate success in a given job. The bulk of our present industrial work does not involve actual trade skill, but rather a limited number of operations which the worker must learn after he is hired and the performance of which depends largely upon his innate *capacity* rather than upon any *proficiency* he has acquired in school or in a previous occupation. It is this type of personnel problem which is most frequently approached by the employment psychologist. The technique of mental tests or measurements of innate capacity is widely used in this connection. There is a further type of personnel problem which necessitates trade tests. This need arises when selecting workers such as carpenters or machinists who profess, at the time they are hired, some trade proficiency. It is desirable to determine by a trade test whether they actually have the proficiency which they claim.

**Fundamental principle of employment psychology.** There is one principle that is absolutely fundamental in dealing with the above problems. The tests or other measurements to be used in selecting persons for a given occupation must be evaluated by giving them to persons whose actual ability in that occupation is known and *efficiency in the tests compared with efficiency in the occupation*. In other words, we must not devise a test that seems plausible, trust that it will work, and start using it for employment purposes. We must first *test the test*. If workmen who are good in the test are good in the occupation and those who are poor in the test are poor in the occupation, then the test is valid, while if there is no consistent relation between occupational ability and test score the test is useless. In the latter case the test is promptly scrapped. In the former case, if the test is given to a prospective employee who has never worked at the occupation in question and he makes a high score, it is fairly safe to predict that he will be successful in the occupation after he has learned it, while if he makes a low test score it is probable that he will be unsuccessful even after long training. The procedure is, of course, not as simple as outlined here and subse-

quent chapters will discuss the methods in considerable detail. However, this principle of testing the tests is central to the whole problem and its observance marks the difference between scientific and unscientific employment psychology.

**Where employment psychology is most valuable.** If an establishment is contemplating the introduction of psychological methods of employment, the question naturally arises as to where it will be most profitable to begin. It is not feasible and probably not worth while to devise psychological methods for every job. It is naturally desirable to place the effort where it will do the most good. This involves two problems: (1) determining where the need is greatest and (2) determining where conditions are such that psychological methods will be valid.

The management usually has a pretty good notion of the locus of the greatest need. A high labor turnover often indicates occupational misfits. Other things, of course, contribute to turnover, but the square peg in the round hole is no mean factor, and it is usually possible to determine whether the other factors are important in a given case. The need for these employment methods depends further on the relation between applicants and vacancies. If the number of applicants for work of a given sort is no greater than the number of vacancies, selective methods are unnecessary because no selection can be made. Every one who applies must be hired. If, however, the applicants exceed the vacancies in number, it is necessary to hire some and reject others. There is then opportunity for psychological methods to aid in the selection of those who have the greatest promise of success.

From the standpoint of psychological technique there are two considerations involved in determining where such methods will be valid. In the first place, the measurements must be standardized on a considerable number of workmen. There is a danger in statistics of basing results on too small a number of observations. A meteorologist would not measure the temperature for two days during one summer in order to predict the temperature the next summer. It would be equally absurd for a psychologist to standardize a mental test on two lathe operators, a good one and a poor one, with a view to predicting the ability of others who were tested. The

oretically one should use a sufficient number of operators so that if any more were added to the group the results would not be significantly changed. Few psychologists would be content with less than twenty individuals and fifty are better. Consequently, if there are only a few persons working at a given job it is useless to try to standardize upon them any occupational tests for that job. The other technical consideration involved is the attitude of the workers on whom the measurements are being standardized. A mental test is worthless unless the person taking it does his best. Tests are designed to measure a person's maximum capacity of a given sort. If one does not exert himself to maximum effort the results are meaningless, for the superiority of one person to another in test score may merely signify that he tried harder and not that he possessed any superior ability. Consequently, if the workers on whom the test is to be standardized are hostile so that they will intentionally do poorly or make no effort to follow directions, or if the proposition cannot be presented to them in such a way that they will take it seriously, it is better not to attempt it at all. This arousal of proper attitude in many cases, however, calls merely for tact on the part of the management and the psychologist. When workers understand the real purpose of the testing program, they will realize that it is being carried out for the advantage of prospective workers as well as of the management. They will appreciate it as a serious matter and will cooperate. Having thus determined where the need for more efficient selection of employees is greatest and where there is a good prospect of valid results, the employment psychologist may then embark upon his program of testing the tests.

**Employment psychology and human welfare.** But the program to be discussed has still wider implications in the social order, and before plunging into details it will be well to consider employment psychology from the broad standpoint of its contribution to human welfare. This is desirable because of the feeling existing in some circles that any methods aimed at increased industrial efficiency are one-sided — benefiting the employer, but not the employee; and further, because of the prevalent impression that such methods treat the worker as a machine and evaluate and dispose of him in

automatic fashion. These notions are without foundation as far as the psychologist is concerned.

Such methods, to be sure, are usually initiated by the management and obviously because they expect them to work to their financial benefit. But this does not mean that the principles adopted will not also benefit the employee. Both employer and employee are tremendously concerned with proper placement of the individual. The occupational misfit is an economic loss both to the company and to himself. He naturally decreases production, but he also decreases his own pay. He has little prospect of advancement, and the ultimate outcome is often his dismissal or his voluntary separation from the concern. He might have spent his time more profitably in learning an occupation for which he was better adapted. It may sometimes seem to be an immediate hardship to refuse a man a job for which he is unqualified, but it is doubtless a kindness to him in the end. Moreover, it is frequently a question, not of rejecting him altogether, but rather of finding some other place in the plant where he will qualify. Economic waste hits all of us including the workman himself, and there is no waste more far-reaching than misdirected human activity. Employment psychology tries to alleviate this misdirected activity by placing individuals in that particular occupation where they stand the greatest chance of success.

Vocational adjustment may proceed from either end. We may take the individual and attempt to determine in which one of many vocations he has the greatest promise of success. This is usually termed "vocational guidance." Or we may take a group of applicants for a job and determine those that are best qualified. This is usually called "vocational selection." The present book is concerned only with the latter. The two fields, however, are not unrelated. As vocational selection develops standards for hiring people for various jobs, those standards can be subsequently used in guiding individuals. If, for instance, tests or other methods have been devised for selecting machinists and salesmen, it will be possible to give both sets to a youngster seeking a vocational objective and tell him in which of these directions he stands the greatest chance of success. It will be many years, of course, before occupa-

tional standards are developed in sufficient numbers to make possible a comprehensive vocational guidance program along these lines, but the results can be used as fast as they become available. There is another way in which the work of selection indirectly contributes to guidance. If a person is refused a job in which the prognosis is very unfavorable, this increases his chances of locating something for which he is better adapted, and if he is prevented from entering a number of vocations in which he would never have a future, he is more liable to land where he belongs. Employment psychology thus works to the interest of the employee as well as of the employer.

The other prevalent notion, that the scientific employment process is automatic and mechanical in character, should be somewhat tempered. To be sure, the techniques to be discussed in the following pages may give this impression to a certain degree. The procedure must in its large outline be rather objective and impersonal, but it is hoped that in many cases it will be supplemented by other factors and by the good judgment of the employment department. For instance, a worker who represents the third generation of the same family which has been employed in the mill constitutes a social factor that cannot be overlooked. An individual who is temporarily inefficient because of some disability should naturally receive special consideration. An applicant whose morale is temporarily disturbed by external factors should be treated as a special case. The notion of the square peg in the round hole is not to be construed literally as an absolute, inelastic proposition. To some extent the man influences the job and the job influences the man, and there are many instances where the fit was originally slightly imperfect, but where minor changes produced a very effective result. The notion has been rather recently advanced of the "worker in his work unit." This unit involves the worker's capacity, interest, and opportunity. The most satisfactory results will come about through the interplay of these three. The worker needs certain minimum capacities in order to stand any chance of success in his job, but he also needs opportunity to develop those capacities and possibly others, and he needs such interest in the work as will enable those capacities to function adequately. It is

even probable that in some instances ability will conflict with interest. The latter should then be treated with respect and it should not be forgotten that the worker is an individual. If he feels that brick-laying is the one sort of work in all the world that interests him, this should receive some consideration in his final placement. Interests, however, are often due to an individual's experience rather than to any innate factor and to that extent are perhaps somewhat less of a fixed entity than are his capacities. A person sometimes likes a given job because he started there originally or because he has friends who induced him to go into that particular work. Or perhaps he prefers to work in New York near the bright lights, although there are better openings elsewhere. If an applicant is manifestly unfitted for a given job, but is tremendously interested in it, this situation calls for tact on the part of the employer in showing him his small chance of advancement in this line and his better possibilities in some other line. Effort may well be made in such cases to interest the applicant in some other kind of work for which he has the requisite ability. Even if he starts out with a definite interest in a given job, but without the ability, it is quite probable that in the course of time, when success has not come, his interest will wane. A common type of interest that leads to considerable confusion in vocational adjustments is the desire possessed by a great many workers for a white-collar job. In some circles there appears to be a certain social stigma attached to an occupation which involves more or less dirt. This stigma is entirely unfounded. The work of the man in overalls is often more of a social contribution than that of the man in the white collar. After all, an individual's greatest contribution is to be made in the line for which he is best fitted. It is better for a man to be an expert machinist than a poor lawyer or to be an efficient carpenter than an ineffective physician.

These problems of vocational adjustment have still wider implications in the social order and extend beyond the machine shop or the stitching room. The maladjusted worker constitutes a serious social problem. He is apt to be in economic difficulty and even in straitened circumstances because, if he is engaged in work for which he is not qualified, he is likely to be penalized in his compensation.

It is probable that this factor contributes materially to poverty and the ills that go with it. It may likewise contribute to even worse things. Many delinquents or criminals can be accounted for through economic failures. Typical of this class is the individual (often of low intelligence) who is hired for one job after another, but fails in each after rather extensive trial, finally becomes discouraged, and either shoulders a tin can and starts up the track or else begins with petty larceny and goes from bad to worse. Being refused at the outset some of these impossible jobs rather than being permitted to waste time trying to master them might have sifted the individual until he reached a place where he could fit. Furthermore, these maladjustments lead to dissatisfaction and unhappiness. A considerable portion of our industrial unrest is due to the fact that workers are not engaged in those types of work for which they are suited. The continuous uphill effort and the subtle feeling of not getting along, while old age and sickness and unforeseen emergencies stand in the offing, give the worker's life an emotional undercurrent that is undesirable. It may express itself in his attitude toward his family or toward his employer or toward his fellow man in general. Having a job for which he is adapted will appreciably alter this undercurrent of dissatisfaction and unhappiness.

So the employment psychologist is confronted with the immediate problem of selecting men for a particular job, but is also indirectly concerned with the more remote but more far-reaching social problem of vocational adjustment. If every factory operative, every office worker, every one on the road and every one at an executive's desk could be doing that type of work for which he was best adapted and in which he was most interested, the world would be a better place. The following pages will discuss psychology's humble contribution to these ends.

Outline. The next chapter discusses the psychological "gold brick." It is desirable to dispose of these pseudo-psychologies before proceeding to discuss scientific methods. There is so much misuse of the term "psychology" and there are so many things on the market purporting to be applied psychology that it seems best to clear the ground at the beginning. Chapter III sketches the history of scientific vocational psychology. Inasmuch as mental



tests play a large rôle in employment psychology they are dealt with in a general way before proceeding to their actual application. Chapter IV describes typical mental tests with which a psychologist should be familiar before engaging in employment research. Chapter V deals with general test technique — the devising and administering of tests.

As suggested above, the tests for a given occupation must be evaluated by comparing test scores with ability in the occupation. This latter factor is technically called the "criterion." The methods of obtaining this criterion — estimates by foremen or production figures — and means for combining various criteria into a single one are discussed in Chapter VI. The next chapter considers the "subjects" or workmen on whom the measurements are to be standardized. The distinction between measurements of capacity and proficiency has already been made. The former of these may be divided into special capacity such as memory, attention, judgment or reaction time, and general capacity or intelligence. Chapters VIII and IX deal with special mental capacities in relation to actual vocational performance. The former discusses the case in which we attempt to devise a special test that reproduces the total mental situation involved in the job and the latter the method of dividing and analyzing the job into its mental components and measuring them separately. The technique of comparing or correlating test and criterion is described and illustrated for various occupations. Chapter X treats general mental capacity or intelligence in somewhat similar fashion. A separate chapter (XI) is devoted to interests. The employment psychologist is beginning to realize that other things besides ability are of importance, particularly a person's interest in and attitude toward his occupation. The scientific work upon these factors is not as far advanced as the work upon capacities, but such results as are available are presented.

There are many aspects of personality that we are at present unable to measure — such things as honesty, tact, leadership. Some information regarding these is often desirable in employment problems. It is at present necessary to depend in such cases on the judgments or estimates of persons who know the applicant in question. However, it is possible to obtain these estimates in

fairly scientific fashion and to make them considerably more reliable than if obtained in the ordinary manner. These methods are discussed in Chapter XII on rating scales. There are various miscellaneous factors more or less related to vocational aptitude that are sometimes used in lieu of, or as a supplement to, tests. Some of these are discussed in Chapter XIII — academic record, personal history blank, letter of application, recommendations, and the interview.

Methods of measuring proficiency as contrasted with capacity, i.e., trade tests, are discussed in Chapter XIV. The technique for devising and standardizing them is given with examples of the different kinds. Chapter XV deals with job analysis and specifications. It does not cover the entire field and method of job analysis, but confines itself mainly to the place of psychology in the more complete program of job analysis. The last chapter deals with the present status and the future possibilities of employment psychology.

## CHAPTER II

### PSYCHOLOGY GOLD BRICKS

#### THE INTELLECTUAL UNDERWORLD

If a man of wealth became interested in psychological methods for analyzing the mental traits of his children or his employees and this interest received due publicity, he would shortly be waited upon by a delegation from the intellectual underworld. Some would propose to read the horoscope of the parties in question, others to study the lines on the palm of the hand, others to feel the bumps on the heads under consideration; another would bring along a neurotic friend who could go into a trance and communicate with some deceased relative in the spirit world to see what *he* thought about it, while still others would present methods for predicting character and future success from the shape of the forehead, ears, nose, or chin, from bodily posture or gait, or even from the position in which the middle vest button was characteristically worn. If our friend asked these various persons individually if their technique was psychological, they would undoubtedly answer in the affirmative. If he inquired whether their methods were infallible or whether they could predict with only a certain margin of error, they would assure him that their results were certain, that they never made mistakes, and that error was a thing with which they were unfamiliar and for which they had no use. If he took the trouble to have a number of them make their observations and predictions separately, he would probably find them contradicting one another on salient points. If, on the other hand, he chose at random and followed the advice of one member of the delegation in planning, for instance, his child's career or in promoting minor executives, he would doubtless find ultimate results markedly at variance with the prediction. Suppose that after this experience our friend read an advertisement for a book like the present one purporting to discuss the application of scientific psychology to the type of problem in which he was interested. His first reaction

could be easily predicted. He could not be blamed for rejecting such a work as another of those "psychological" things with which he had had such unfortunate experience. He could not be expected to discriminate between real psychology and pseudo-psychology because he had never studied the former and no scientist had ever enlightened him. That is why in the present discussion of employment psychology it seems best to clear the ground at the outset; before telling what psychology can do, to tell what it does not do, and to mention some things that have no right to masquerade, as at present, under the name of psychology.

The reason for its existence. The reason for the existence of this pseudo-psychology is obvious — it pays the one who is promoting it. It lends itself admirably to advertising and to commercial exploitation. The "prospect" is confronted with statements and proposals about his own mind. These intimate matters naturally arouse his interest and arrest his attention, and this is a first step in the sale. Furthermore, in the mind of the layman, there is a certain atmosphere of mystery surrounding "psychology." There is a natural credulity toward the unknown or little understood. This makes one somewhat prone to believe the indefinite statements of the "applied psychologist," and belief is a second important step in the sale. It has thus been possible to capitalize interest and credulity and lead persons to accept a proposition, disguised with pseudo-psychological terminology, which they would reject under other circumstances. Consequently, the last few years have witnessed a mushroom crop of "applied psychologists" who never saw a laboratory or clinic, but who pay big income taxes; an avalanche of literature about obtaining health, happiness, and success by the use of various "systems" of vibrations or mental dynamism which contain no more psychology than the solar system; the development of institutions of learning which teach "divine metaphysics" and other subjects related to mental efficiency; and lately the advent of the "mental broadcasting station" which broadcasts treatment or advice to subscribers.

Main objection of psychologists. The real psychologists object to this sort of thing primarily because it is presented under the guise of psychology. If it were called "galomalism" or some other

meaningless name and the gullible wished to invest, it would be nobody's business. But it is called "psychology," and when the promised improvement in memory fails to materialize, when the blonde employees fail to come up to expectations, when the position of the vest button proves to be non-differential of the salesman's success, and when the inspirational phonograph records and the psychology hymns fail to raise the salary, then real psychology gets the blame.

**Present extent.** The number of such gold bricks that are on the market is appalling. There are popular "psychology" magazines filled with encouragement and inspiration for those who are unfortunate or ambitious. A large amount of advertising space is devoted to courses and systems. In many of the large cities applied psychology clubs have been organized to meet periodically for exercises in concentration. When a salesman comes to town with a supply of gold bricks he is given excellent publicity by the press and when he leaves he takes thousands of dollars with him. There could be no better vindication of the late Mr. Barnum's famous epigram.

It is beyond the scope of the present work to discuss all aspects of pseudo-psychology. Consideration will be given only to those which, at one time or another, have purported to have a bearing on vocational or employment problems. Critical discussion of mental efficiency methods and therapeutic devices must be omitted. Regarding these only one suggestion will be made. If a person claims to be a "psychologist," it will be found illuminating to ascertain if he is a member of the American Psychological Association. This is the official organization of scientific psychologists. It has rigid requirements for membership, and if a person is elected it is highly probable that he is a real psychologist. If he is not a member, it is pertinent to inquire the reason. Very few persons outside the Association know enough psychology to be using it to any great extent in a practical way. The Association publishes an annual directory of its members and at almost any university or college some member of the psychology department will be glad to answer inquiries as to whether certain names appear in that directory.

## ASTROLOGY

Astrology is one of the oldest methods of analyzing character, but it is still with us. Its absurdity may be best shown by a quotation from a recent guide to character analysis. Persons born in February

are very intuitive and good judges of character and human nature. They are successful in mercantile interests and enterprises. It is said that the best wives are born in this month being always faithful and devoted. Great sincerity and power are possible for those born in this month. They rise to great heights and on the other hand are inclined to sink to the lowest depths. . . . Their most common diseases are of the nervous and rheumatic orders. They should guard their actions on the ninth and sixteenth day of each month. . . . They will excel in music and art and should marry those born in October, January or June. (246, 8.)<sup>1</sup>

This quotation may seem too absurd to be worth mentioning, but statements similar to the foregoing were recently radiocast from one of the Eastern stations and horoscopes read over the radio for persons who submitted the requisite personal data. In writings on astrology the absence of statistics is noticeable and there is apparently no effort to ascertain empirically if the alleged relations exist.

Astrology, however, has actually appeared in the employment office. The writer knows the man who does practically all the hiring for an industrial concern with a personnel of over a thousand. There are certain types of very dusty work which many who are hired are unable to stand. The employment man's theory is that persons who are born in the spring are unsuited for this job. This theory is doubtless the result of some observation. The date of birth is one of the items on the application blank and contract, and it is natural that the employment man should occasionally notice it at the time an employee was leaving on account of the dust. The writer inquired whether record was ever kept of the birthdays of all who left under these circumstances to ascertain if any were not born in the spring, but this had never been done. A few days later the writer was called in to get a conclusive "proof" of the theory by interviewing a man who was leaving for the above reason and who

<sup>1</sup> Throughout the book figures in parentheses refer to the bibliography. A figure in italics indicates the page in the reference denoted by the preceding figure.

admitted that he was born in April. Similar cases of astrology in the employment office could doubtless be found. At any rate, it would be well for a chief executive to ascertain whether anything of this sort is operating in his own concern. Many a man who is normal in most respects may make some utterly illogical generalization like the foregoing on the basis of a few observations. It is a thing of which no record would be made, but which might very considerably color personal opinion in judging applicants.

#### SPIRITUALISM

Spiritualism likewise plays a greater rôle in vocational and employment problems than is realized. Great numbers of persons attend séances seeking information and advice, and many of the questions asked fall within this field. Many of the teachers of a high school near Boston regularly consult a famous medium in the vicinity. A member of the New York Stock Exchange whose name is familiar to most of the readers of this book regularly consults a certain medium before going to the Exchange or embarking on any important venture. Many localities have a practitioner who by spiritualism or something similar solves problems for a host of clients. Such individuals presumably do not work in the employment office, but have been known in some instances to advise on special personnel problems such as promotion or transfer in executive positions, as well as to give individuals advice in selecting the type of work for which to apply.

Of course an occasional visit to a spiritualistic meeting is sufficient to convince the scientifically minded of the inanity of the whole affair. Very general suggestions are put forth, such as "the spirit of an elderly lady with gray hair and a dark dress," or a dim spot of illumination appears in the dark room and whoever identifies the individual in question gets the message. Distressed relatives may catch the least word which remotely indicates that the spirit which they seek is in communication with them. One little sign which appeals to their waiting imagination shatters ordinary caution and emotion supplants reason. There are some instances in which men of note whose sons have died have become champions of the spiritualistic cause. The writer himself has received mes-

sages from friends in the spirit world, but he always had to help the medium or reader along a little in order to get the message.

**Futility unless telepathy can be demonstrated.** While the psychological researchers have collected a quantity of data purporting to be authentic communications from famous people who have died (Plato sometimes writes English, by the way), there is one theoretical point to consider. There is no presumption that a mind in this world can communicate with a mind in the spirit world unless it can be proved that two minds in this world can communicate without some physical medium such as light or sound. Telepathy has not been demonstrated under carefully controlled laboratory conditions.

**Laboratory experiments on telepathy.** A few careful researches of this sort have been carried out by psychologists and their results are of more value than hundreds of random observations or unscientific experiments. An experiment conducted in the Harvard Psychological Laboratory used a sort of split choice reaction. (619.) In the ordinary reaction experiment a person sees one of two lights appear, and if it is the one on the left he presses a telegraph key with the left hand, and if it is the one on the right, he presses a key with the right hand. In this experiment one person, the "agent," observed the lights and another person, the "percipient," tried to operate the keys appropriately by "reading the mind" of the first person. The agent had before his eyes a small box with a hood attached. At the back of this box was a point of light to serve as a fixation mark. On each side of this were two small square areas either of which could be illuminated electrically. According to which was illuminated the agent concentrated on the right or the left. The percipient had his hand on a little platform mounted very delicately so it could be moved to the right or left with practically no effort. The task of the percipient was to move this platform or key in the direction of which the agent was thinking. An automatic shuffler presented the lights in a perfectly random order. A buzzer signaled to the percipient by two faint sounds when a light had been presented to the agent. He then tried to think whether the agent was concentrating right or left and to move his key accordingly. The exposure continued for 30 seconds and just be-



fore the end of this time a circuit was automatically closed which recorded the position of the percipient's key as right or left. There was then a 50-seconds rest before the next trial. The entire apparatus worked automatically and the number of correct and incorrect reactions was recorded on electric counters. The control and recording apparatus was in a separate room. The agent and percipient sat in Morris chairs in a sound-proof room about six feet apart in complete darkness. Under the experimental conditions according to the laws of chance it was to be expected that 50 per cent of the reactions would be correct, just as in tossing a coin a large number of times heads will be thrown approximately half the time. In other words, chance or mere guessing would account for approximately 50 per cent correct answers. A significantly greater number than this would indicate the presence of some other factor — possibly telepathy. As a matter of fact, the result of about 600 observations yielded 47 per cent correct answers.

Some rather extensive telepathy experiments were conducted at Leland Stanford University. (133.) A few typical results will be cited. Numbered blocks (numbers from 20 to 99) were drawn from a bag by the agent and effort made to transmit them to the percipient in 1000 trials. The probability of getting the tens digit correct by accident was 12.5 per cent and the actual number correct 10.2 per cent. Moreover, the agent sometimes thought about the number by simply holding a visual picture of it vividly in mind, sometimes by imagining the muscular sensation of speaking the number, and sometimes by imagining the sound of some one else speaking it. The percipient likewise recorded whether the numbers came to him in visual terms or in terms of the feeling of the speech muscles or of sound. There was no correspondence between the way the agent thought of it and the way the percipient received it.

A card-guessing experiment was conducted with about one hundred different agents and percipients, the agent in each instance drawing a card and concentrating on it and the percipient trying to tell what it was. In determining the color of the card, accident would account for 50 per cent correct answers with the actual results 49.8 per cent correct. As to the number on the card, chance would give 10 per cent correct and the actual results were 10.5

per cent. As to the suit of the card with a chance expectation of 25 per cent, the result was 26 per cent. There was no appreciable difference in results whether the agent and percipient were 1 or 10 yards apart or whether the time interval for each trial was 10 or 60 seconds. Moreover, the experiment was tried with 10 percipients who were "psychics," i.e., claimed that they had ability along these lines of telepathy and clairvoyance. Their results were not substantially different from those of the other percipients.

Other scientific experiments of this sort might be cited, but the foregoing are sufficient to make the point that telepathy has not been demonstrated under carefully controlled laboratory conditions, and until this is done it is useless to consider the possibilities of a mind in this world communicating with a mind in another world. Persons who accept advice from the spirit world for vocational or other purposes are flying in the face of science and putting themselves at the mercy of ignorance or unscrupulousness in the form of a medium.

#### PHRENOLOGY

Phrenology is another type of pseudo-psychology that is still current. A New England concern a few years ago engaged a phrenologist to work in its employment office. The writer was on one occasion himself mistaken for a phrenologist. When it became noised about the office and factory that a psychologist was to begin work, a number of persons, it was discovered later, expected to have the contour of their skulls examined.

Semblance of scientific basis. Phrenology did have historically a little more semblance of a scientific basis than the other pseudo-psychologies mentioned above. Science had discovered that certain parts of the brain were concerned with certain sensory or motor functions. If a portion of the skull was removed and the surface of the brain stimulated, movements of certain muscles might take place, and by stimulating different parts of the brain different muscle groups could be made to contract. Moreover, injury to a certain portion of the brain often left a person with some defect such as inability to see or hear or speak. Now when real scientists were presented with these facts, they set out to analyze the matter fur-

ther by experimenting on the brains of living animals, by post-mortem examination of the brains of persons who during life had some mental or motor defect, and by dissection and microscopic examination to trace neural pathways from the sense organs and muscles to their destination in the brain. It was slow work and it is not yet completed. But all the phrenologist needed was a good start afforded by the knowledge that there was at least some brain localization no matter of how coarse a variety. It seemed plausible enough that if there was a brain center for movement of the arms there should likewise be centers for memory, reverence, combativeness, conscientiousness, constructiveness, etc. Scientific method was too slow and laborious for the phrenologist. He made casual observations of his acquaintances, noting a little cranial protuberance here and there and attempting to find some mental trait of the individual to correspond, but neglected to ascertain whether any people with a similar protuberance lacked the trait or whether any with the trait lacked the protuberance. Thus he built up a system and mapped out the skull in an utterly illogical and unscientific fashion. This movement started about 1800 and there has been very little revision of the principles originally laid down. A work written in 1832 is still the standard to-day!

**Assumptions of phrenology.** There are at least three assumptions made by phrenology that are erroneous. In the first place, it assumes that there are a great number of specific traits or faculties that have their function located in a particular portion of the brain. All the evidence of scientific experiment, however, shows that the brain does not function in as small units as those claimed. It has been possible to locate regions concerned with various muscle groups and with vision, hearing, and most of the other senses. But no detailed areas have been found to be concerned with such things as high *versus* low tones or sensations of red *versus* blue. A map of the functional areas of the brain made by the scientist is very simple compared with that made by the phrenologist. Moreover, there are some parts of the brain with which no very definite function has as yet been found to be correlated, but phrenology long ago mapped the entire surface. A notion of the discrepancy between the actual findings of science and the assumptions of phrenology may be ob-

tained by a detailed consideration of a few regions of the brain. The first column in Table I lists the functions of certain brain regions that have actually been determined, while the second column gives the corresponding functions assigned to those regions by the phrenologist. Attention is especially called to the location of "rever-

TABLE I. SCIENTIFIC VS. PHRENOLOGICAL STATEMENTS AS TO FUNCTION OF CERTAIN REGIONS OF THE BRAIN

ACTUAL FUNCTION AS DETERMINED BY EXPERIMENT	FUNCTION ALLEGED BY PHRENOLOGISTS
Movement of feet and legs Movement of trunk and shoulders Movement of hand and fingers Movement of jaws and lips Auditory sensations Touch, temperature, and muscle sensations Visual sensations Maintenance of equilibrium	Reverence Marvelousness Ideality Constructiveness Destructiveness Hope Love of children Amativeness

ence" in the region actually concerned with the movement of the feet and legs and of "amativeness" in the region actually concerned with the maintenance of equilibrium. Furthermore, the phrenologist locates memory in the front lower central part of the brain, a region the function of which has not as yet been scientifically ascertained. It is actually found that the localization of memory follows that of the sense department involved — injury to the visual region of the brain causing disturbance of memory for visual details, but not for auditory details. The functions in the left column include practically all whose location the scientist considers determined. The phrenologist, on the other hand, presents a map comprising the functions of thirty-five different regions.

The second erroneous assumption of phrenology is that there is a direct and obvious relation between the development of a trait and the size of the corresponding region of the brain. There is, to be sure, evidence of a slight relation between the size of the brain and intelligence, but the complexity of structure is equally important. When it comes to the development of the small regions with which the phrenologist is concerned, the difference, if any, would be prac-

tically invisible. For instance, it is pretty well established that speech is controlled by an area on the left side of the brain. Microscopic work indicates that the layer of gray matter in the corresponding region of the right side is not quite as thick, but the difference in thickness is not over a millimeter. No phrenological methods could detect a difference of this magnitude.

The final assumption is that a few casual observations afford a sufficient basis for generalization. This is, of course, contrary to all scientific method which implies the collection and statistical treatment of large numbers of observations before drawing conclusions. Books on phrenology comprise an analysis of a relatively small number of individual cases rather than a statistical treatment of large numbers of persons. The absurdities to which this technique has led are manifest in Table I. So while phrenology had a more plausible basis than the other pseudo-psychologies, its fundamental assumptions are unsound and it has absolutely no contribution to make to scientific employment psychology.

#### PHYSIOGNOMY

Physiognomy is probably the most widely used of these questionable methods of analyzing character or predicting mental capacity. If it is construed in a wide sense to include the appearance of the face and head and entire body, it will be found quite widespread. Many firms require a photograph with the application blank in instances where the person is not available for an interview, or use the photograph to select those who are to be interviewed. In some types of work attractive personal appearance is, of course, a requisite, or race may be significant; but there is often a feeling that something of further value may be obtained from observing the photograph. Probably some aspect of the features influences the judgment, perhaps unconsciously, of the one evaluating the application. The head of a large technical school arranges his interviews with the boys who apply for entrance in such a way that they have to walk down a long aisle before reaching his desk. He believes that he obtains valuable insight into their traits or capacities by observing their gait during their approach. One employment man has an antipathy to red hair. An office manager eschews

blondes in his organization. It is then worth while to consider scientifically the value of such methods for employment or vocational purposes.

Popular belief in physiognomy is doubtless back of its more practical use and commercial exploitation. In our literature and in our personal contacts we have been taught to attach significance to the shifting eye, the high forehead, the receding chin, the dimple, the heavy jaw, the short neck, and even the erect posture or the shambling gait. These beliefs have developed like many of our other unscientific notions as a result of casual observation combined with an absence of logic. Then, when we come to consider character from the practical standpoint of employment, we merely carry over uncritically these notions that have developed in the popular mind. The basis of these popular beliefs may now be analyzed in a little more detail.

Association by similarity. It is a fundamental law in psychology that one thing is apt to suggest or call to mind another which is similar to it. Thus the photograph of a friend suggests that friend because the contours of the former are similar to those of the latter. "Robin" may suggest "oriole"; or "cat" may suggest "tiger," for the same reason. This principle of analogy or association by similarity operates in our popular notions about physiognomy. A person with a short neck suggests a bull and we then attribute to him some of the stubborn characteristics of that animal. Cats are crafty and treacherous and clams are cool, flabby, and inert, and hence arises the importance which we attach to the feline tread or the clammy handshake. By the same principle of similarity a broad forehead suggests a broad mind, hard-textured flesh suggests a hard heart, and sharp features a sharp, penetrating intellect. Or, again, if the physiognomy of a stranger is like that of an acquaintance it is quite natural to attribute to the former the traits of the latter. If one has had a disagreeable personal experience with some one whose hair is red, he may assume that another person with similar hair is likewise irascible. These popular generalizations, then, are readily explainable by the law of association, but this does not justify them. The fact that one thing reminds you of something else does not establish it as a scientific truth that there is any real

relation between the two. The popular mind is content with its assumption, and if sometimes the relation proves subsequently to hold and sometimes the reverse, it is customary to remember the former instances and to forget the latter.

**Observation influenced by expectation.** Another principle which is involved in the development of popular physiognomic notions is that we tend to see what we expect to see. If our attention is set for some particular aspect of an object, it is that part which we see first or which impresses us most vividly. In a familiar laboratory experiment in which a pointer swings along a scale and a bell rings at some particular point, if an observer is attending to or thinking about the bell he will judge that it sounds at an earlier position of the pointer than he will otherwise. Attending to the bell facilitates its entrance into consciousness. Or, again, if one attends to the trombone in an orchestra he can hear it stand out from the other instruments. A motor mechanic will detect a main bearing knock that the layman would overlook because the mechanic takes an attitude of expectation. This principle then operates to substantiate our beliefs in physiognomy. If a person shakes hands weakly we expect that he is going to show some vacillation, and while he perhaps manifests that trait no more than do other persons with whom we come in contact we are all "set" for it in his case and notice instances which would otherwise escape us. Or, if we observe some one with large ears and have been taught that these denote parsimoniousness, we watch for instances which might be construed as manifesting that trait and magnify them, although our friends with small ears may be acting in a similar manner. But once we observe these expected traits they serve further to confirm our generalization, as another case "which proves it."

**Evidence of habitual activity.** There are some aspects of popular physiognomy, however, which seem to have an objective reason to account for them instead of being dependent purely on the association process or attention attitude of the person making or corroborating the generalization. It seems plausible at first glance that certain habitual activities should leave their impression in observable form on the face or body. A studious person bending over his books for years may become round-shouldered. A pugilist may

develop a tendency to look at his adversaries — and every one else — with his head turned toward the left and bent slightly forward. A philosopher may contract his brows while he ponders until the wrinkle becomes permanent. A criminal may repeatedly avoid the gaze of his prospective victims till his eye becomes “shifty.” While it is perfectly true that certain habitual tendencies may affect the musculature in a permanent fashion, there is a fallacy involved when it comes to reversing the proposition and assuming that those with round shoulders are studious, that those with a sidewise gaze are belligerent, that those with wrinkled brows are philosophers, and that those with unsteady eyes are criminalistic. Suppose, to take a more obvious example, that all the Chinese students in a university (probably by reason of language handicap) scored less than 78 points in an intelligence test, it would be obviously fallacious to argue that all students who scored less than 78 points were Chinese. Yet this is exactly the same type of logical fallacy that is committed in assuming that the round-shouldered are studious, etc. As a matter of fact, there are other things that might equally well cause round shoulders, such as crap-shooting; or that might produce an unusual position of the head, such as rheumatism or infantile paralysis; or that might wrinkle the forehead, such as near-sightedness; or that might cause the eye to shift, such as shell shock. Popular beliefs regarding physiognomy have then no scientific basis. They are used, however, by many persons in practical problems of predicting human characteristics and this uncritical use must obviously lead to many mistakes. Moreover, our popular notions pave the way for our acceptance of systems of character analysis that have been commercialized.

**Commercial systems of physiognomy.** It was quite natural that the astute purveyor of psychology gold bricks should avail himself (or herself) of the fertile field of physiognomy in which the seeds of popular belief were already sprouted. If persons had some notions regarding the relation between the face or figure and character, why not devise a detailed system — arbitrary, to be sure, and without scientific foundation — and sell it to them? This is precisely what was done. The promoters wrote books and articles, and gave lectures and, best of all, personal consultation and advice, using such criteria as the following:



Texture is a great classifier of humanity. The individual of fine hair, fine-textured skin, delicately chiseled features, slender, graceful body and limbs, as a general rule is refined, loves beauty and grace, and likes work either purely mental in nature or offering an opportunity to handle fine delicate material and tools. On the other hand, the man with coarse hair, coarse-textured skin, and large, strongly formed features inclines as a general rule to occupations in which strength, vigor, virility, and ability to live and work in the midst of harsh, rough, and unbeautiful conditions are prime requirements. . . . Blondes as a general rule are changeable, variety-loving, optimistic and speculative, while brunettes are consistent, steady, dependable, serious, and conservative. . . . The man who resembles a greyhound in form is quicker, keener, more responsive, and less enduring than the man who resembles the bulldog in form. . . . Poets, educators, and essayists will show a marked tendency to resemble the triangle in structure of head and body — both head and body wide above and narrower in the lower portions. Generals, pioneers, builders, engineers, explorers, athletes, automobile racers, aeronauts, and others who lead a life of great activity, will show a general tendency toward structure on the lines of the square — square face, square body, square hands. Judges, financiers, organizers, and commercial kings will show a general tendency toward structure upon the lines of the circle — round face, rounded body, and a tendency to roundness in the hands and limbs. (246, 39.)

Some of their work is merely a more literary restatement of popular beliefs and some of it is dogmatic assertion.

#### EXPERIMENTAL EVALUATION OF CHARACTER ANALYSIS FROM PHYSIOGNOMY

While the theoretical basis of such generalizations seems unsound and while the criteria are mere observations and not actual measurements of the physiognomic characteristics in question, the crucial point is to determine experimentally whether the alleged relations actually exist. Suppose photographs are available of a group of intimate acquaintances who can give a pretty reliable estimate of one another. It is possible then to obtain a notion as to a person's status in each of a number of mental traits, that status being the combined judgment of his acquaintances. The photographs may then be submitted to judges who have never seen the individuals in question and they may be required to estimate each person in each trait from his photograph. Then these estimates made from physiognomy may be compared with the actual traits

as indicated by the combined judgment of acquaintances to determine whether physiognomy under these conditions has any validity in indicating the mental traits. This may best be done by the procedure of "Correlation." This is a statistical method for indicating the closeness with which any two variables or sets of traits or measurements are related. If, for instance, those who are rated by their acquaintances as most intelligent are likewise rated from the photographs as most intelligent, and *vice versa*, we speak of a "high positive correlation." If those actually most intelligent are judged from the photographs to be least intelligent, and *vice versa*, we speak of a "high negative correlation," while if there is no tendency one way or the other we speak of a "zero correlation" or "no correlation." By the use of proper formulæ<sup>1</sup> it is possible to determine a "correlation coefficient" which indicates, not merely whether the correlation is high, low, or negative, but exactly how close is the relation between the two variables. A coefficient of 1.00 indicates perfect correlation — i.e., the person who is highest in one variable is correspondingly high in the other, the person who is next highest in one is proportionately high in the other, and so on down the list. From 1.00 the coefficient can range down through zero to -1.00, which indicates a perfect negative correlation. In actual practice a coefficient less than .30 does not attract much attention, while with a coefficient of .50 there is manifestly some relation between the things that are correlated, but there is still considerable error in trying to predict one thing from the other. A further notion as to the meaning of correlations of a different magnitude may be obtained from the following consideration. Children of the same family resemble one another to some extent in physical characteristics. Twins resemble one another more strikingly in these respects. In some instances such physical characteristics have been measured in pairs of children from the same family and correlation coefficients computed between various physical characteristics of a

<sup>1</sup> Appendix I illustrates the computation of such coefficients and gives a notion of the significance of correlations of different magnitudes. In the examples presented there, sets of scores in test and job are given. These are then ranked. In the present connection the original estimates on the basis of physiognomy or acquaintance consist of ranks so that the computation would begin with the third and fourth columns in the examples.

child and those same characteristics of the brother or sister. These correlation coefficients prove to be somewhere around .40. On the other hand, when a similar procedure is followed for a group of twins the correlations are somewhere around .80. Consequently, when in the course of the following discussion we find a correlation of about .40, we may think of the two things in question being related to about the extent that brothers and sisters resemble one another in physical characteristics; while if we come across a coefficient of about .80, we may think of the two things in question being related as closely as twins resemble one another.

Estimates of miscellaneous traits from physiognomy. A few studies of the sort just mentioned may be cited. A group of 25 college women rated one another in a considerable number of fairly definite traits. (244, 37.) Each individual took the names of the 24 others and considering, for instance, "neatness," selected the one she considered neatest of all and marked her 1, then selected the next neatest and marked her 2, etc., so that the 24 were arranged in rank order from the neatest to the least neat. Then the same thing was done for refinement, sociability, and a series of other traits, each one being rated separately and each woman ranking all the other 24 women. There was then available for each woman 24 estimates of her possession of a trait; e.g., she had been assigned a ranking in neatness by all the other women. These 24 figures were then averaged to get the consensus of opinion of the entire group regarding that particular woman's neatness. Similar averages were found for her refinement, sociability, etc. This procedure was repeated for each woman. This combined judgment of 24 acquaintances might be taken as about the best statement of the real characteristics of the women that could be secured. Having then obtained these figures, photographs of the 25 women of uniform style and size were submitted to a group of men who were totally unacquainted with the women involved. Each man ranked the individuals with reference to neatness as far as he could judge it from the photographs, marking the neatest 1, the next neatest 2, etc. Then he ranked them with reference to refinement, with reference to sociability, etc., making his estimates entirely on the basis of the photographs inasmuch as he did not know the individuals at all.

It was then possible to compare or correlate the ranks assigned by any one man on the basis of the photographs with the ranks assigned by the combined judgment of acquaintances. In exactly the same way the photographs were submitted to another group of women totally unacquainted with the original group and they ranked them as the men had done on the basis of the photographs.

To consider the matter first from the standpoint of the combined estimates of the judges rather than from that of the accuracy of the individual judge, the ranks assigned to any photograph for a given trait by all the men were averaged. These average estimates from photographs were correlated with the combined judgments of acquaintances above mentioned. The same was done with the women's estimates from photographs. The results are shown in Table II. The first column gives the traits involved; the next

TABLE II. CORRELATION BETWEEN AVERAGE ESTIMATES OF TRAITS FROM PHOTOGRAPHS AND AVERAGE ESTIMATES OF THOSE SAME TRAITS MADE BY ACQUAINTANCES <sup>1</sup>

TRAITS	ESTIMATES OF PHOTOGRAPHS BY 25 MEN	ESTIMATES OF PHOTOGRAPHS BY 25 WOMEN
Neatness.....	.03	.07
Conceit.....	.10	.27
Sociability.....	.29	.29
Humor.....	.21	.45
Likeability.....	.30	.45
Intelligence.....	.42	.61
Refinement.....	.50	.52
Beauty.....	.60	.49
Snobbishness.....	.58	.53
Vulgarity.....	.61	.69
Average.....	.36	.44

<sup>1</sup> From Hollingworth's *Judging Human Character*, by permission of D. Appleton and Company, New York.

column gives the results when the group of men are estimating the traits from photographs, and the last column gives the results when the group of women are using the photographs as a basis for judgment. For instance, the combined opinion of acquaintances regarding the neatness of the individuals in question correlates with

the combined opinion of a group of men (based only on photographs) regarding the neatness of the same individuals to the extent of .03. The former combined opinion correlates likewise with the combined opinion of a group of women (based only on photographs) regarding the neatness of these same individuals to the extent of .07. Similar figures follow for the other traits.

Remembering that 1.00 represents perfect correlation and .00 none at all, it is obvious that estimates from the photographs are none too satisfactory for practical purposes. Moreover, the value of the estimate seems to depend on the trait. Vulgarity, snobbishness, and beauty seem to be estimated fairly well from the photograph, while neatness, conceit, and sociability are quite the reverse. One would hesitate to use physiognomic diagnosis of many of the traits indicated for employment purposes even if he could obtain 25 judges to make the physiognomic estimates.

Although the results are none too satisfactory when the estimates from photographs made by a group of 25 judges are pooled, the situation is much worse if we consider the validity of an individual judge. In the usual employment situation there are, at most, only a very few persons who evaluate a given applicant from his physiognomy. Instead of using the average estimates from photographs as in Table II, we may take the estimates made by one judge from the photographs, with reference to neatness, for example, and correlate these estimates with the combined estimates of the acquaintances regarding neatness. To indicate the typical trend, 10 judges are taken at random and their individual correlations for three of the traits given in Table III. The estimates of intelligence made, for instance, by Judge A using the photographs, correlate with the combined opinion of acquaintances regarding the intelligence of the same individuals to the extent of .51. The estimates of sociability made by Judge C correlate with the combined opinion of acquaintances regarding sociability to the extent of .05.

Inspection of the table shows a big variation between judges. Some of them estimate a trait from physiognomy fairly well and others rather poorly. For instance judge A estimates intelligence with a correlation of .51, while Judge D actually has a negative correlation; i.e., to some extent tends to place those actually of high

TABLE III. CORRELATION BETWEEN ESTIMATES OF TRAITS FROM PHOTOGRAPHS MADE BY INDIVIDUAL JUDGES AND AVERAGE ESTIMATES OF THOSE SAME TRAITS MADE BY ACQUAINTANCES <sup>1</sup>

JUDGE	INTELLIGENCE	NEATNESS	SOCIABILITY
A	.51	.11	.39
B	.11	.10	.08
C	.15	.29	.05
D	-.27	.06	.49
E	.08	.24	.08
F	.43	.41	.28
G	.04	.11	.02
H	.39	-.09	.32
I	.22	-.08	.00
J	.30	.02	.55
Average.....	.20	.12	.23

<sup>1</sup> From Hollingworth's *Judging Human Character*, by permission of D. Appleton and Company, New York.

intelligence toward the low end of the scale and *vice versa* when he considers their photographs. In neatness the best judge is C with a correlation of .29 and the worst is H with a correlation of  $-.09$ . In sociability J is the best (.55) and I the worst (.00). Moreover, a judge who estimates one trait well may fail when another trait is involved. Judge A, for instance, is fairly competent to estimate intelligence (.51), but manifestly incompetent to estimate neatness (.11); J estimates sociability with some validity (.55), but his estimates of neatness have no validity at all (.02). Consequently, it would seem hazardous to attach much practical significance to physiognomic estimates of this sort made by one or at most a few individuals. All-round judges of character from physiognomy are apparently scarce.

Estimates of intelligence from physiognomy. It is possible to make a more careful check than in the foregoing instances with reference to estimates of intelligence from physiognomy because they can be compared with intelligence as objectively measured by tests, whereas physiognomic estimates of other traits such as neatness or sociability must be evaluated by comparison with judgments of acquaintances. In one such study (10), 63 managers, buyers, and assistants in a large department store were given an intelligence

test somewhat similar to that used in the army (*infra*). Their photographs were then submitted to 12 graduate students interested in these personnel problems. The student judges were required to estimate the intelligence of these business men from their photographs. After looking through the pictures to get a general idea, each judge selected the 7 most intelligent and 7 least intelligent, then the 14 who were superior but not as good as the first 7, and likewise the 14 who were inferior but not as poor as the lowest 7. Arbitrary values were assigned to each of the four classes in order to handle the data statistically. The 12 estimates of the intelligence of a given manager were then averaged to get the combined opinion of the judges regarding his intelligence. A similar average was obtained for each of the other men concerned. These combined estimates of intelligence from photographs were then correlated with the actual intelligence as measured by the tests. The correlation coefficient is only .27. Moreover, it must be remembered that these results used only the extremes of intelligence and did not include the middle group. Had this been included the correlation would probably have been smaller still.<sup>1</sup> It would seem that even when a dozen persons pool their results estimates of intelligence from physiognomy are almost worthless.

The same experiment may be considered from the standpoint of the validity of the individual judge. One of the simplest methods is to note for each judge how many men he places on the right side of the average and how many on the wrong side; i.e., whether a man he rates in the best 7 or superior 14 is actually above the average in measured intelligence or not. These results are shown in Table IV. Judge A, for instance, places 27 individuals correctly on the basis of these photographs; i.e., if they are actually above the average in tested intelligence, he places them above from physiognomy and *vice versa*. However, he misplaces 15 individuals; i.e., judges them as above average when they are actually below or *vice versa*. There are only two or three judges who place many more of the men correctly than incorrectly and some actually have more in the incorrect column. The total for the correct column is only 17 per cent more than that for the incorrect column. So while

<sup>1</sup> For statistical reasons beyond the scope of the present work.

TABLE IV. NUMBER OF PERSONS CORRECTLY PLACED AS ABOVE OR BELOW AVERAGE INTELLIGENCE BY INDIVIDUAL JUDGES ON THE BASIS OF PHOTOGRAPHS<sup>1</sup>

JUDGE	NUMBER OF PHOTOGRAPHS ON CORRECT SIDE OF AVERAGE	NUMBER OF PHOTOGRAPHS ON WRONG SIDE OF AVERAGE
A	27	15
B	23	19
C	25	17
D	23	19
E	19	21
F	26	16
G	22	20
H	22	20
I	20	22
J	20	22
K	22	20
L	22	20
Total.....	271	231

<sup>1</sup> After Anderson.

pooled judgments are bad enough, individual judgments are worse, and it would manifestly be useless for one or two persons to use such physiognomic methods for practical purposes.

One other study may be cited. Ten photographs of college freshmen, rather uniform in pose and in mounting, were ranked according to apparent intelligence and this order correlated with intelligence as actually measured by one of the standard tests (325). There were 108 different judges who made such estimates and whose individual results were correlated with actual intelligence. The distribution of their correlation coefficients is given in Table V.

The table indicates, for instance, that with 10 of the judges the correlation of their estimate with actual intelligence is between  $-.39$  and  $-.20$ , while 22 of the judges yield correlations between  $-.19$  and 0. It is to be noted that there are relatively few of the correlations over  $.40$ , and hence there is no very great indication of the ability of individuals to judge intelligence from photographs. One further problem was approached in this study, namely, having a group of judges sit together as a committee of from two to six members, and give a combined committee judgment rather than an individual judgment. The results of the committee correlations



TABLE V. CORRELATIONS OF MEASURED INTELLIGENCE WITH ESTIMATES OF INTELLIGENCE BASED ON PHOTOGRAPHS, FOR DIFFERENT JUDGES<sup>1</sup>

CORRELATIONS	NUMBER OF JUDGES GIVING CORRELATIONS INDICATED	NUMBER OF "COMMITTEES" GIVING CORRELATIONS INDICATED
-.74 to -.40		4
-.39 to -.20	10	8
-.19 to 0	22	17
0 to .19	37	14
.20 to .39	14	9
.40 to .59	15	
.60 to .79	8	
.80 to .89	2	

<sup>1</sup> After Laird.

are given in the third column of Table V. There are four instances where a committee yields a correlation between  $-.74$  and  $-.40$ , eight instances between  $-.39$  and  $-.20$ , etc. Obviously, the judges are no more effective when operating as a committee than when operating alone. This all goes to indicate still further the difficulty of estimating such a thing as intelligence from physiognomy.

**The halo effect.** One other point that has rather wide implications in the whole theory of rating procedure may be noted in these experiments on judging miscellaneous traits from photographs. It is brought out by correlating estimates of various traits with one another to determine, for instance, whether persons who are rated high in humor are likewise rated high in perseverance, kindness, etc. Photographs of twenty women were ranked by judges with reference to six traits, and the average rank of each individual obtained in each trait. (246, 46.) Using these average ranks each trait was then correlated with each of the others. The results are shown in Table VI. Any figure in the table indicates the correlation between the trait listed at the left of that row and the trait listed at the top of that column. For instance, the correlation of humor and intelligence is  $.47$ , that of perseverance and humor  $.33$ , etc. It will be seen that humor, perseverance, kindness, courage, and intelligence all seem rather closely related. A person who looks as if he possessed a high degree of one of these appears as if he possessed a high degree of the others. Conceit and deceitfulness,

TABLE VI. CORRELATIONS BETWEEN ESTIMATES OF DIFFERENT TRAITS MADE ON THE BASIS OF PHOTOGRAPHS<sup>1</sup>

TRAITS	INTELLIGENCE	HUMOR	PERSEVERANCE	KINDNESS	CONCEIT	COURAGE
HUMOR.....	.47					
Perseverance.....	.88	.33				
Kindliness.....	.76	.65	.39			
Conceit.....	.28	-.03	.08	-.56		
Courage.....	.89	.43	.79	.72	-.25	
Deceitfulness.....	-.11	-.28	-.03	-.69	.66	-.49

<sup>1</sup> From Hollingworth's *Vocational Psychology*, by permission of D. Appleton and Company, New York.

on the other hand, show negative or low correlations with these former traits, but correlate highly with one another. These results suggest a danger in ratings of this sort — a factor that plays a rôle in ratings in general. There seems to be a tendency for the judge to form a general impression that is favorable or otherwise and then to rate the person accordingly in a number of traits. This effect has been called a “halo effect.” This halo of general impression often colors estimates of various traits so that not much validity can be attached to the estimate of any one trait *versus* another. The judge thinks he is evaluating the traits independently, but he is merely recording repeatedly his general impression.

The results of such studies as the foregoing are not encouraging to those who hope to predict character from physiognomy. When estimates of mental characteristics made from photographs are compared with more certain criteria of those characteristics, such as the judgment of intimate acquaintances or measurements of intelligence, there are marked discrepancies between the two. An individual judge's results have little validity, and even when a considerable number of judges pool their estimates the results are far from what is to be desired. The only conditions under which it would be at all advisable to install such methods for employment would be where a corps of probably twenty or more persons were available to make these physiognomic judgments and average their findings. It is a question whether this procedure would be expedient. Inasmuch as scientific methods are available that do not

necessitate the use of such a corps, it would seem wiser to devote one's effort to the use of such scientific methods. These will be described in later chapters.

**Evaluation of commercial systems of character analysis.** Psychologists have been so busy improving their methods of using mental tests and other measurements for practical purposes of employment that they have devoted little effort to experimental refutation of specific relations between aspects of physiognomy and mental characteristics that are assumed by commercial systems of character analysis. A few investigations of this sort, however, have been made and the results are presumably typical of what will be found when further alleged relations are studied.

**Alleged blonde and brunette traits.** One of the most widely known systems of character analysis makes much of the mental differences between blondes and brunettes. As this is an easily observable anatomical distinction, it would be very convenient if we could infer character therefrom. According to the system in question, this is possible, and a list is provided of the traits possessed primarily by blondes and a similar list is furnished for the brunettes. It was possible statistically to determine the validity of these lists. (441.) Twelve "blonde traits," such as positive, dynamic, driving, aggressive, domineering, and fourteen "brunette traits," such as negative, static, conservative, were arranged in a random order on a printed blank. These blanks were given to 94 persons who were above average intelligence. Each person selected two pronounced blondes and likewise two pronounced brunettes with whom he was very well acquainted. For each of these acquaintances he went through the printed list of 26 traits and marked them with a plus or minus sign according to whether, in his judgment, the person possessed that trait or not. The persons marking the blanks were not familiar with the particular system of character analysis involved and the traits occurred in a random order so that the alleged blonde ones would not be found grouped together. It was then possible to tabulate the per cent of blondes who were rated plus on the blonde traits and also who were rated plus on the brunette traits. These results are shown in Table VII. For instance, 81 per cent of the blondes are positive, which is an

TABLE VII. PER CENT OF BLONDES AND BRUNETTES RATED AS POSSESSING ALLEGED BLONDE OR BRUNETTE TRAITS<sup>1</sup>

BLONDE TRAITS	187 BLONDES	187 BRUNETTES
Positive.....	81	84
Dynamic.....	63	64
Driving.....	49	50
Aggressive.....	62	56
Domineering.....	36	36
Impatient.....	56	51
Active.....	88	82
Quick.....	70	68
Hopeful.....	85	85
Speculative.....	53	51
Changeable.....	53	43
Variety-loving.....	66	62
<b>Average.....</b>	<b>63</b>	<b>61</b>
BRUNETTE TRAITS		
Negative.....	16	17
Static.....	28	31
Conservative.....	51	61
Imitative.....	39	40
Submissive.....	25	26
Cautious.....	54	60
Painstaking.....	56	61
Patient.....	43	52
Plodding.....	27	31
Slow.....	20	24
Deliberate.....	47	57
Serious.....	58	72
Thoughtful.....	67	70
Specializing.....	52	45
<b>Average.....</b>	<b>42</b>	<b>46</b>

<sup>1</sup> After Paterson and Ludgate.

alleged blonde trait, but 84 per cent of the brunettes are likewise positive; 63 per cent of the blondes are dynamic, but so are 64 per cent of the brunettes. While brunettes are supposed to be negative, 17 per cent are found to be so in the actual results, but 16 per cent of the blondes are likewise. The averages indicate that the 12 alleged blonde traits are possessed in general by 63 per cent of the blondes, but are likewise possessed by 61 per cent of the bru-

nettes; the 14 alleged brunette traits are possessed on the average by 46 per cent of the brunettes, but also by 42 per cent of the blondes.

A somewhat different approach was made to this same problem by sending to 50 well-known sales executives a list of traits used in the system of character analysis referred to above. (278, 244.) Each executive selected four highly successful salesmen and checked on this list of traits the ones they possessed. Results were available for 152 salesmen. The outstanding characteristics mentioned were: positive, dynamic, driving, aggressive, active, quick, painstaking, hopeful, patient, serious, thoughtful, specializing. Of these seven are "blonde" and five are "brunette" traits. Obviously it would be difficult to select a good salesman from his complexion.

One other bit of evidence bears on this same question. The system alleges that persons of mechanical bent are typically of light complexion. In a survey of 400 metal workers, most of whom were presumably of somewhat mechanical bent, 16 per cent were light, 32 per cent dark, and 52 per cent medium. (292.) Obviously there is no tendency for them to be typically light. The majority are medium, and there are more dark than light complexions in the group.

**Miscellaneous physiognomic factors.** From current systems of character analysis a considerable number of miscellaneous physiognomic characteristics were selected which were claimed to be an index of mental traits and these physiognomic characteristics were actually measured. (128.) Persons intimately acquainted with the individuals who were measured provided estimates as to these particular mental traits. Furthermore, the individuals were placed on the stage before a group of judges who were unacquainted with them and they were estimated casually for the mental traits to determine the possibility of a practitioner by "intuition" being able to estimate traits in an interview, although he actually professed some physiognomic basis for his judgments. The traits studied were the following: judgment, intelligence, frankness, ability to make friends, will-power, leadership, originality, and impulsiveness. These traits were selected because there was fair agreement among the physiognomists regarding them. The physical measurements were made with calipers, sliding compass, steel tape, and head-

square. The character analysts use only their eyes. These experimenters used instruments which must have made their measurements at the worst far more accurate than the character analysts' at their best. They measured then a large number of physical characteristics which the analysts claimed correlated with the mental traits above enumerated. There were anywhere from 20 to 36 different items measured in connection with each of the eight mental traits, making a total of 201 different measurements obtained upon each individual.

Thirty students were measured in this fashion and were rated as to the mental traits by other members of their fraternities or sororities. These ratings prove to be quite reliable; i.e., the different members of the fraternity or sorority agree rather closely with one another in rating a given individual. These opinions of acquaintances thus form a pretty good standard by which to evaluate the physiognomic measurements. On the other hand, the reliability of the physiognomic measurements is low. For instance, there are a number of them that are alleged to indicate judgment. If the relative standing of the students in one such set of measurements is obtained and correlated with their standing in another measurement which is supposed to indicate the same mental trait, the correlations are uniformly small. In other words, the theories of the character analysts with reference to physiognomic indications of a given mental trait are discordant among themselves.

The crucial point is, of course, the correspondence between the physiognomic measurements and the estimates made by close associates. The best way to summarize the entire results is to average all the correlations of the physiognomic measurements for a given trait with the associates' judgments of that trait. For instance, with intelligence there were 29 different factors measured. Each of these is correlated with estimated intelligence. The average of these 29 correlations is then computed and, as indicated in Table VIII, gives .03. Similar averages for the other mental traits appear in the first column of the table. It is obvious that these correlations are all extremely small and show practically no relation as between the alleged physiognomic indicators of mental traits and the actual possession of those traits. The average correlations between

TABLE VIII. CORRELATION BETWEEN RATINGS OF CLOSE ASSOCIATES AND PHYSIOGNOMIC FACTORS<sup>1</sup>

TRAITS	CLOSE ASSOCIATES AND PHYSIOGNOMIC MEASURES	CASUAL OBSERVERS AND PHYSIOGNOMIC MEASURES	CLOSE ASSOCIATES AND CASUAL OBSERVERS
Judgment.....	-.01	.14	.32
Intelligence.....	.03	.05	.02
Frankness.....	.05	.15	.21
Friendliness.....	-.11	.19	.18
Will-power.....	-.07	.04	.26
Leadership.....	-.04	.07	.31
Originality.....	.09	.08	.32
Impulsiveness.....	.10	-.07	.20

<sup>1</sup> After Cleston and Knight.

the opinion of the casual observers who interviewed the subjects and the physiognomic measurements are given in the next column and are likewise of insignificant magnitude. The results for the close associates and for the casual observers are correlated in the last column. These correlation coefficients are slightly higher than the others and might indicate a very slight possibility that the judges, through "intuition" or something of the sort, are able to evaluate certain aspects of personality. However, only three of the traits yield correlations as large as .30 and the other five are distinctly less. The general conclusion of the study is that "the average of 201 correlations between various physical traits purported to reveal variations in character traits and our criterion is .00 with the correlation varying from .00 as chance would account for. Physical measurements which underlie character analysis agree neither with themselves nor with other measures of character."

**Present extent of physiognomic methods.** It is difficult to ascertain to what extent methods like the foregoing are being seriously used for employment purposes. There is no doubt that many persons are using some popular or personal generalizations of this sort as a supplement perhaps to other criteria. A questionnaire was circulated in 1922 among one hundred employment managers and insurance agency managers asking if they used any system of character analysis, and if so what one. (309.) Sixty-five replies were received—22 from insurance men and 43 from industrial

concerns. Two of the former and four of the latter stated that they were using some system of character analysis. It is probably safe to say that six out of a hundred were using some system rather than six out of 65, because those using one would be more apt to reply than those not using one. Six per cent is not a large figure, but it is six per cent too many to be using this sort of method.

In the light of the experiments on character analysis by means of physiognomy, one wonders why the methods should at present be in use at all — why the system has not already killed itself. The answer lies in the fact that some practitioners are able occasionally to hit the mark and make successful predictions or give valuable advice — ostensibly by the use of their system, but in reality on some other basis. In the first place, the analyst may hit by chance *one of the many* occupations for which the individual is fitted. It is not always a case of there being one job and only one in the world in which an individual may be successful, but there are usually many lines in which he will achieve success. Consequently, selecting one of these by accident is not such a remote possibility. We often do this very thing ourselves, otherwise most of us would be maladjusted. The analyst can often by casual observation eliminate some possible lines of work for which the person obviously is disqualified and thus stand a greater chance of accidental success in predicting from the remainder. In the second place, the analyst may in the course of conversation discover likes and dislikes which may be of some vocational significance. He will perhaps be enabled with this information to make common-sense suggestions quite apart from any system. In the third place, if a person pays for vocational advice and the “expert” recommends a certain line of work, the individual will perhaps try harder than he would otherwise and hence reach a higher level of success than, with his ability, he would ordinarily attain. The “expert,” of course, gets the credit for this. Finally, when persons are discussing such cases and comparing notes they are apt to slip into a very common human fallacy of stressing the cases of coincidence and forgetting the other cases. This same tendency to neglect the negative instances plays into the hands of the pseudo-psychologists. Persons remember the one case in which they hit the mark and forget the other ninety-



nine times in which they miss. Scientific employment psychology may not always hit the mark, but it does so at least far more frequently than does pseudo-psychology.

#### SUMMARY

Before proceeding to the discussion of psychological methods in employment, it is necessary to clear the ground of a considerable amount of pseudo-psychology which is being widely commercialized and is masquerading under the name of psychology to the detriment of the real science. A number of these pseudo-psychologies have played a rôle in employment problems in recent years. *Astrology* has no scientific basis and its generalizations have not been evaluated statistically, but it is actually in use. *Spiritualism* has certainly nothing to contribute until its actual existence can be proven. It is illogical to assume communication with spirits until telepathy can be demonstrated and this has not as yet been accomplished under laboratory conditions. Yet spiritualistic mediums are consulted on various problems of a vocational nature. *Phrenology* started with the scientific findings regarding the functions of certain regions of the brain, but went far beyond the experimental results. It erroneously assumed a much more detailed localization of functions and a direct relation between the functional capacity of a brain region and its size, and it used a few casual observations as a basis for generalization.

*Physiognomy* is the most prevalent of these pseudo-psychologies. Our popular beliefs in it are due largely to the fact that one thing we see is associated with similar things (e.g., a short neck suggesting a bull and hence stubbornness); to the fact that our observations are influenced considerably by what we expect to see (e.g., a weak hand-shake causing us to watch for further indications of vacillation); and to our assumption that, inasmuch as habitual activities often leave bodily traces (e.g., the round shoulders of the studious), it is logical to argue backward from those traces to the activity in question. These popular beliefs, however, pave the way for our acceptance of commercial systems of character analysis from physiognomy. The validity of such beliefs and systems has been to some extent studied scientifically. Estimates of various mental traits made on the basis of photographs by judges who never saw

the original individuals were compared with careful estimates of those same traits by intimate acquaintances or with actual measurements of the traits. The results indicate that a single judge is very inaccurate in making such estimates from physiognomy, and that while matters may be improved somewhat by using a considerable number of judges and averaging their results, the correspondence of this pooled estimate and actual possession of the trait is not sufficiently close to make the physiognomic factor of much practical value. Moreover, the results are often vitiated by the "halo" effect or tendency to get a general impression of good or bad and rate the person high in most desirable traits, or *vice versa*, instead of evaluating the traits independently.

These studies show the futility of judgment of character traits from physiognomy when the judge is left to his own devices. The futility has been shown to be equally great when it is a question of the relation between specific physiognomic and mental characteristics claimed by commercial systems of analysis. The alleged relation of complexion to specific character traits is without foundation, for it has been shown statistically that blondes possess the traits that are supposed to characterize brunettes to just as great an extent as do the brunettes themselves, while the brunettes rival the blondes in the possession of the alleged "blonde traits." A group of intimately acquainted persons rated one another in several traits which have received considerable attention from the character analysts. Alleged physiognomic correlates of these traits were actually measured with calipers and steel tape — 200 physical measurements upon each individual. These measures were separately correlated with the criterion provided by estimates of intimate acquaintances. The correspondence between the physiognomic measures and the actual traits is exactly what would have been expected by chance. The physical measurements give no indication whatever of the mental traits in question.

The practical employment man is certain sooner or later to come in contact with some of these pseudo-psychologies especially with the commercial systems of character analysis. From the foregoing considerations it is obviously to his interest to confine his efforts to scientific employment psychology rather than to invest in any of these psychology gold bricks.

## CHAPTER III

### HISTORY OF SCIENTIFIC VOCATIONAL PSYCHOLOGY

THE previous chapter called attention to some of the pitfalls of pseudo-psychology which beset the practical man to the detriment of himself and of his attitude toward the real science. The perspective in which we view employment psychology may be still further enlarged by a consideration of its historical background.

#### INDIVIDUAL DIFFERENCES

**Early interest in general laws.** As mentioned in the introduction, the early studies in psychology were directed toward the determination of general laws, whereas the differences between individuals are usually much more significant for practical purposes. Aristotle developed laws of association to explain why one idea calls up another; Weber and Fechner worked out the psycho-physical law to express the relation between the intensity of the stimulus, such as light or sound, and the intensity of the sensation; Ebbinghaus derived certain laws pertaining to the memory. While this type of work was of immense importance in laying the foundations of the theoretical science, it had little to do with sorting the applicants for a job. It was only after the fundamental groundwork had been partly laid and some psychologists turned their efforts from the general principles to the individual differences that progress was made in the field which is our present concern. A number of factors contributed to this shift of interest.

**More detailed study of faculties.** The earlier psychology made a good deal of the notion of "faculties" into which mind could be divided, such as the faculty of memory, or the faculty of attention. It became evident, however, that these faculties must be still further subdivided. It developed that memory for numbers and memory for words were two quite different things and that attention to a thunderclap and attention to an uninteresting book differed. In order to investigate these more minute differences it was necessary to arrange appropriate experimental material. Lists of

words and lists of numbers were devised for the study of those two types of memory. Interesting and uninteresting materials were selected for the study of the different kinds of attention. This sort of material was the prototype of the mental test. Then, when the experiments were conducted to determine the difference between memory for words and memory for numbers, it became obvious that, while such differences existed, there were also differences between individuals in their ability to retain the material. In the experiments on attention, while the expected difference between attention to interesting and uninteresting material was obtained, there also proved to be differences between persons in the amount to which they could adequately attend. In this way the early experimental psychologists, attempting to subdivide such "faculties" as memory and attention, noted and became interested in these individual differences.

Need for mental measurements in problems of heredity and education. On the other hand, practical considerations came from without the science to meet halfway the interest that was developing within. Galton and others were much interested in heredity. It was observed that many students who took honors at Oxford had parents who had done likewise; that one family had many lawyers and judges, while another had musicians and artists among its ancestors; that some individuals with phenomenal memory had parents who excelled in that same respect. A good deal of data of this sort was collected, using qualitative estimates of the traits or abilities in question. Students of these problems came to realize the need for quantitative data and for some method of actually measuring the traits. Hence they looked to the psychologists for assistance in devising such measurements. Education was another field which early realized the need for psychology. It was observed that one child made rapid progress in school, while another was retarded. One individual sixteen years of age might be entering college, while another of the same age was still in the fourth grade. What caused this difference in educational performance was a moot question. It led some of the pioneers to seek methods for measuring general ability or whatever mental factor was involved in school retardation.

Thus the interest of theoretical psychology itself in devising finer measurements for the study of general abilities, and the interest of those in other fields, such as heredity and education, in obtaining such measurements to use in their own problems, led to a considerable shift in emphasis from the general laws to the individual differences.

#### EARLY DEVELOPMENT OF MENTAL TESTS

**Freshman mental tests at Columbia.** The outstanding pioneer effort in the use of mental tests was at Columbia University in 1894. Under the direction of Cattell there was instituted a plan for testing the students in their first and fourth years. The purpose of the project is well expressed in the following paragraph which actually constituted the material for one of the memory tests:

Tests such as we are now making are of value both for the advancement of science and for the information of the student who is tested. It is of importance for science to learn how people differ and on what factors these differences depend. If we can disentangle the complex influences of heredity and environment we may be able to apply our knowledge to guide human development. Then it is well for each of us to know in what way he differs from others. We may thus in some cases correct defects and develop aptitudes which we might otherwise neglect. (114.)

The tests used were for the most part those of sensory capacity, such as color blindness, auditory acuity, perception of pitch, sensitivity to pain, or else measurements of the speed and accuracy with which certain tasks could be accomplished, such as marking 100 letters or making 100 movements. This project is typical of the early test work. Miscellaneous tests were devised and tried in order to see whether they differentiated persons from one another and in order to determine how a given person scored with reference to the rest of the group.

Coöperation in the standardization of tests was the next step historically. After various workers had devoted considerable independent effort to devising miscellaneous tests and trying them out on small groups of individuals that were available, it became obvious that coöperative effort would facilitate matters. Some of this took place, of course, through the publication in scientific journals

of descriptions of material and methods for tests so that they could be tried by other investigators with comparable results. In 1906 the American Psychological Association appointed a permanent committee to act as a general control committee on the subject of measurements. It was charged among other things with the development of a series of group and individual tests. The committee functioned for several years and issued reports upon tests for auditory acuity, pitch discrimination, imagery, and a set of "association tests," involving such things as color naming, cancelling numbers, learning a code, giving opposites, and following complicated and confusing directions. These "association tests" have been widely used since that time in their original form and have likewise served as a pattern for investigators who devised other tests along similar lines. (674.)

Binet. Another significant contribution to the development of tests was the work of Binet. His problem was to devise means of measuring intelligence of children. The method consisted essentially of finding a set of questions for children of each age such that the average child could answer them satisfactorily. Consequently, if a child was backward he would fail on the questions for his own age, although he might succeed in answering questions designed for some lower age. Binet began his work about 1900 and published his original intelligence scale in 1908. (49.) It has subsequently been translated and revised by Goddard, Terman, and others, and is now, in these revised forms, one of the most widely used intelligence tests.

Whipple in 1910 published the first edition of his *Manual of Mental and Physical Tests*. (662.) This presented most of the important tests that had been devised and used to any great extent up to that time. They were classified, standards given as far as available, and considerable data presented on the relation of the tests to each other and to various factors such as age. This compilation of material and procedure for giving a considerable number of tests has been very valuable, as it has enabled many persons to give similar tests under similar conditions and to compare results.

This perhaps marks the high spot in the early development of tests for their own sake. The emphasis up to this point was largely

upon devising tests and standardizing them on various groups of individuals. It was eminently desirable that the technique should go through these stages before being put into the practical situation. Efforts to use tests for hiring employees in 1894 would have been premature. Much had to be learned about the principles to observe in the construction of test material, in the wording of directions, in the selection of time limits, and in the scoring of results. In short, the whole theoretical technique had to be reasonably well developed before it was profitable to apply the tests to practical ends.

#### COMPARISON OF TEST SCORES WITH OCCUPATIONAL ABILITY

The next step in the history of employment psychology was the comparison of efficiency in the tests with efficiency in the occupation. If those who were effective in the occupation made high test scores and *vice versa*, the tests could then be used with applicants for a position to predict their future ability therein. The pioneer efforts in this field were made by Münsterberg about 1911 with his study of motormen of the Boston Elevated Railroad. (408.) His test, to be described in another connection, consisted essentially of an endless belt arranged to pass under a small opening so that the person being tested might discriminate between different figures in different locations on this moving belt and react to them according to their significance. The novel feature was that the test was given to actual motormen and the test scores compared with their service records. It developed that motormen with a good record and few or no accidents made somewhat higher scores in the test than did those with a bad record of accidents.

Münsterberg also gave a series of tests to girls in a school for telephone operators. The tests themselves involved such abilities as memory for numbers, judgment of distances, rapidity of movements, and speed of association. The progress of the girls in the school was compared with their test scores and some tendency was manifest for those with satisfactory progress in learning the operations of a telephone operator to make higher scores in the tests than did those with unsatisfactory progress. The advance made in these studies was fundamental. Hitherto the tests had been standardized on anybody. Now they were standardized on persons en-

gaged in a particular occupation and efficiency in the tests compared with efficiency in the occupation. This same procedure is in use at present, namely, testing the tests. Statistical methods have improved, ingenuity in devising tests has increased, and many technical points have been perfected, but the general principle is the same.

Shortly after this time various other psychologists began to compare test scores with occupational criteria in similar fashion. Scott started his work on methods for selecting salesmen, comparing test scores with sales records. (521.) At Carnegie Institute of Technology there was organized the Bureau of Salesmanship Research which embarked on a five-year program of coöperative research along these lines.

Then came the war.

#### PSYCHOLOGY IN THE WAR

**Organization of psychologists.** On April 6, 1917, it happened that a group of experimental psychologists from the eastern part of the country were gathered at Harvard for one of the informal conferences such as they often have, to talk over their problems. News came of our entry into the war sometime during that forenoon, and after luncheon Professor Yerkes, who was then president of the American Psychological Association, called the group together. He stated that the country was now at war, that other scientists would be putting their knowledge and technique at the disposal of the Government, and that it was likewise the duty of the psychologists to make whatever practical contribution they could to the military situation. Every one conceded that the human factor was an important one in the army and navy and even among those who remained behind. Those who had prior to that time hesitated to launch into the field of applied psychology, feeling that the time was not yet ripe and that the theoretical basis had not been sufficiently laid, immediately cast aside that hesitation in the face of the national emergency. The afternoon was then spent in discussion of psychological war problems which it might be profitable to investigate and of methods of approaching those problems. This was the first war conference of American psychologists. Many of the group



went home afterward and immediately began planning specific methods for dealing with the problems suggested.

A few days later the officers of the American Psychological Association met and started the organization of the psychologists of the country. The military problems were classified as far as possible and committees and subcommittees appointed. These were at first rather informal, with no official status other than that given by the Association. Later, however, they were reorganized as subcommittees of the psychology committee of the National Research Council. They worked upon a wide variety of military psychological problems, such as the psychological examination of recruits, selection of aviators, gun-pointing, night-observing, training and discipline, incapacity, reeducation, emotional stability, self-control, propaganda, and tests for deception.

This is not the place to recount the work of all of these committees. Some were dealing with problems analogous to the employment problems of industry and some with entirely different problems. Only the former will be discussed at all in the present connection. Suffice it that the war advanced applied psychology at least ten years in a few months. When it became absolutely necessary to do something, we found that there was much more psychology to apply than we had realized.

**General mental examination of recruits.** One of the committees above mentioned took up the problem of general mental examination or intelligence testing of recruits. It seemed plausible that different branches of the service and different ranks might require different degrees of general intellectual capacity. Accordingly, a group of psychologists who had previously been most closely in touch with problems of intelligence measurement undertook to devise a test that could be given to large numbers simultaneously, which could be scored by clerks who did not have psychological training and which would yield a reliable indication of general mental ability. Prior to this time most testing had been individual, i.e., one person at a time was examined verbally by a skilled examiner. It would obviously have been impossible for the available skilled examiners to examine individually a million men. Starting with the best available information regarding Binet and allied tests,

a preliminary form was devised and tried out on a few thousand men. In the light of the results it was revised and developed into its final form, the "Army Alpha" test. It was ultimately given to 1,726,000 men. Its uses were many and varied, but some of them were similar to those encountered in current employment psychology.

For instance, there was the problem of eliminating from positions of responsibility those of such low mental status as to render them dangerous. In the army some 8000 men were discharged because mentally unfit for duty; another 10,000 were assigned to labor battalions because of their low intelligence and 9000 were sent to developmental battalions for observation.

Then there were problems of promotion, recommendation for officers' training camps, and the like. As a matter of fact it was found that in the army the average intelligence of the commissioned officers was higher than that of the non-commissioned officers, while these in turn excelled the enlisted men in average intelligence. This fact might be used in promoting men from the ranks. If the officers had higher intelligence, it seemed plausible that a private of high intelligence, other things being equal, constituted better officer material than did a private of low intelligence. Or, again, a few of the best and a few of the worst privates in a company from the standpoint of the commanding officer were selected. This was done for a large number of companies. Then the intelligence of the worst privates was compared with that of the best privates and the former averaged strikingly lower than the latter. These methods of comparing efficiency in the test with efficiency in the job were directly on the main road of progress in employment psychology.

Aside from giving psychologists further experience in employing tests in this way the Army Alpha has since the war been very useful because the items had all been carefully selected after experimentation and standards were available based on nearly 2,000,000 men. Many subsequent experimenters used this test in its army form. Many others have modified it somewhat by way of abbreviation or rearrangement of items. It served as the prototype for most of the *post-bellum* group tests of intelligence.

Selection of aviators. It was realized that ability to fly an air-

plane was rather complex and that the medical examination did not give the whole story, for some recruits who passed all the physiological tests were a failure as aviators. This appeared to be a type of work which required various special mental and motor capacities. Accordingly it was analyzed as far as possible and tests devised for these special capacities, such as speed of reaction, judgment of distance and velocity, ability to detect slight changes in equilibrium and emotional stability. A large number of tests of this sort were given cadets at aviation ground schools. These cadets were followed up at the flying schools to determine their actual ability in flying. These flying results were then compared with the test scores to determine which were the most valuable types of test. Then a smaller number of tests, selected as a result of these preliminary studies, was given to a group of aviators at Kelly Field and test scores correlated with ratings by flying instructors and with the number of hours of flying with an instructor necessary before the man was permitted to fly alone. It was then possible to select a group of tests which would give a fairly good indication of aptitude for flying. The official installation of these methods was under way at the time of the armistice.

The especial contribution of this work to the advance of employment psychology was the greater refinement of statistical methods. In addition to determining the relative importance of the different tests by correlation procedure, they were "weighted." This involved ascertaining just how much importance should be attached to each separate test in the final combined score in order to get the best possible prediction of flying ability. It was found that certain tests overlapped, i.e., to some extent measured the same thing. For example, separate tests were devised for attention and speed of reaction, but the two things were not necessarily independent. A man with good attention tended to be a little quicker in reacting because he paid closer attention while waiting for the signal. This came out when the two tests were correlated, and it developed that those who were quickest in reacting to some extent excelled in attention and *vice versa*. Hence the attention and speed of reaction tests were spoken of as overlapping. In similar fashion it was found that those who excelled in a memory test like-

wise were superior in an attention test, presumably because paying better attention facilitated learning the memory material. Here again the two tests overlapped and correlated appreciably with each other. It was then necessary to make allowance for this overlapping of tests — otherwise one particular trait such as attention might receive undue importance or “weight” in the combined score. This allowance was made by partial correlation — finding the extent to which a given test correlated with flying ability when the effect of the overlapping tests was statistically eliminated. This technique will be discussed later (Chapter VIII). It had been developed previously, but this was one of its first applications in a practical vocational problem.

**Soldier's qualification card.** The committee on classification of personnel dealt with a group of very specific employment problems. Something like half the men in the army have to ply some special trade, and it was obviously advantageous to assign to any such duty some man who already had ability in the trade involved. The problem was then to discover these tradesmen in the draft and make them available. One method of approach was to obtain information in the preliminary interview with the recruit, systematize it, and incorporate it in some standard form. Study of this problem led to the soldier's qualification card. The recruit was interviewed with reference to his personal history, occupational experience, education, etc. These items were entered on a standard card using standard terms and symbols. These cards were then tabbed at the top, the position of the tab indicating the trade with which the man was familiar and the color of the tab indicating his proficiency as far as it could be ascertained. When men of a certain type were needed, it was then possible to look through the files of a unit and select in a few moments, by following down the tabs in a certain position, the men in that unit who were proficient in the trade in question.

**Trade tests.** The above procedure had one drawback. In the interview a man would often make false claims as to his ability in some trade. As a matter of fact it developed that about thirty per cent of those who claimed trade ability were totally inexperienced in the trade in question. If a man was assigned to duty in-

volving carpentry on the basis of his own statement and could not drive a nail, the efforts of the interviewer, clerks, and others were wasted. Hence the trade test was developed actually to measure the man's trade ability — whether he was an expert carpenter or a journeyman or apprentice or whether he was a mere novice. Sometimes a small sample of the job was standardized to test a man's skill, and every item of his performance scored. Sometimes standard questions were asked about tools and materials and processes involved in the trade — questions such as an experienced tradesman should be able to answer. These trade tests were evaluated by comparing scores with actual known ability in the trade. The standards were obtained in various industrial plants using men who had an actual trade status.

The trade test opened up another aspect of employment psychology. Hitherto most of the efforts had been devoted to predicting the aptitude or potential *capacity* of a workman to be successful in some particular job after due training. The trade test measured *proficiency* which the applicant possessed at the time of the test rather than any future possibilities. While the tests of capacity doubtless play the larger rôle in industry, the trade test has its place. The army was the first situation where it was developed on any considerable scale.

**Rating scales.** Army personnel work led to the consideration of certain mental traits, such as leadership, character, or general value to the service that could not be objectively measured. This problem was particularly important in dealing with officers. In the past a certain evaluation of such factors had been made, of course, in considering cases of promotion. But one officer in considering his subordinates would often use entirely different standards of judgment from those used by another officer and would attach different importance to different traits. The committee consequently found it desirable to develop a systematic rating scale covering certain specific qualities. It was ascertained as the result of careful study and evaluation of questionnaires that a limited number of qualities or traits were outstanding in the successful officer. These qualities were carefully defined and their relative importance ascertained. A scale was then arranged with the maximum and

minimum value to be assigned to any quality fixed. The details of the scale will be described in Chapter XII, but it consisted essentially of selecting from well-known officers the names of a few individuals who possessed a given trait, such as leadership, in high, low, or average degree and assigning them standard values in a "master scale." The subordinates who were being rated were compared with this "master scale" making man-to-man comparisons and assigning to the subordinate the numerical value attached to the officer on the master scale most similar to him. This officers' rating scale was one of the first attempts at systematic development of a technique for estimating scientifically these non-measurable mental characteristics.

It is obvious that psychological methods underwent a considerable development during the war. Psychologists are, of course, inclined to magnify the importance of their contribution. The foregoing discussion is introduced, however, not to show how valuable psychology may have proved itself to the country, but how scientific progress was stimulated by the emergency. And inasmuch as many of the problems undertaken were of the type epitomized by "the right man in the right job," this work played an important part in the history of employment psychology.

#### EMPLOYMENT PSYCHOLOGY SINCE THE WAR

**Individual projects.** The termination of the war found psychologists more interested in personnel problems than they had been hitherto. Consequently, many individual research projects were launched in this general field. There was a considerable amount of effort devoted to the development of further tests and the application of them to employment situations. Some psychologists made studies of particular occupational groups as occasion arose. Others went into employment departments as members of the staff — one in a munitions plant, another in a rubber tire factory, another in a silk mill, another in a department store, and several into offices employing large staffs of clerical workers.

**Coöperative projects.** In addition to such individual projects there were other undertakings of a coöperative nature. In several instances a group of business men contributed to a common fund

that was expended by a staff of scientists on some aspect of business research. A company was organized by some of the war psychologists and engaged in consulting work on personnel problems. The National Research Council continued the interest in this field that it had fostered during the war emergency. The Bureau of Public Personnel Administration and the Personnel Research Federation were organized to serve as clearing-houses for information regarding personnel research and to conduct further research. The Psychological Corporation was established in order to thwart the pseudo-psychologist, to promote contact between the person needing practical psychological work and the psychologist best qualified to do that work, to develop various standards and procedures of its own and to devote its surplus to the furtherance of research in psychology. These agencies will be described more at length in Chapter XVI.

The work since the war has been for the most part along the lines that were started during the emergency. Much of the data obtained in the army work has been subsequently analyzed further and critically evaluated. Great numbers and varieties of tests have been devised. Statistical methods have been refined. A large variety of occupations has been studied comparing test scores with some criterion of ability in the occupation. Progress must of necessity be slow and painstaking, but the fundamental principles governing satisfactory work in this field have been pretty well determined. Employment psychology is at present a well-established branch of the science.

#### SUMMARY

The early interest of psychology was in general laws. The shift to the consideration of individual differences came about through theoretical interest in analyzing various mental factors in more detail and through the need of those working in other fields, such as heredity and education, for a technique of mental measurement. The first extensive testing program occurred at Columbia in 1894. Subsequently, there was coöperation in developing and standardizing a variety of tests. A distinct contribution was made to the methods of measuring general intelligence by Binet and the rapidly

growing body of tests for special capacities was collated by Whipple. The next step after the development of tests for their own sake consisted in comparing individual efficiency in tests with efficiency in an occupation. Münsterberg was the pioneer in this field with his experiments on motormen and telephone operators. Just before the war there were several other testing projects under way. During the war the psychologists experimented upon many problems of a vocational sort. The general mental examination of recruits left with us a good group test of intelligence which has been the prototype for many subsequent scales. It also taught us something about the occupational significance of general intelligence. The work of selecting potential aviators gave us insight into the statistical possibilities of weighting a group of tests in order to predict vocational ability. The various qualification cards and blanks devised for army personnel work have been useful patterns for subsequent personnel blanks. The trade test methods called attention to a new field — the measurement of proficiency as contrasted with capacity. The rating scale technique gave us a method of obtaining quantitative data regarding traits that are not directly measurable. Since the war the interest of psychologists in employment problems has continued. Some have worked in employment offices and some have engaged in individual projects using the information and the technique developed during the war. Many coöperative undertakings are now in progress. The application of psychological technique to the selection of employees has gained a permanent place in the increasingly important movement to consider the human element in business.



## CHAPTER IV

### TYPES OF MENTAL TESTS

LIKE any technician the employment psychologist cannot do his work satisfactorily without tools, and the mental test constitutes his most frequently used instrument. He would be as loath to hazard a diagnosis of the mentality of a prospective employee without administering tests as would be the physician to diagnose bronchitis without the use of a stethoscope. The psychologist has occasion to remark, "Work as fast as you can without mistakes," about as frequently as the physician has to repeat, "Cough, and say 'ah.'"

The subject<sup>1</sup> is sometimes tested orally, sometimes with blanks on which he writes or marks, sometimes with objects such as puzzles and sometimes with simple mechanical contrivances which he manipulates. In all cases, however, the aim is to measure some capacity or proficiency in order to predict what the individual will do at some future time and under certain circumstances — for instance, when learning a particular job. It is not possible, of course, to measure the entire capacity in question any more than it is possible for a manufacturer to evaluate carefully every pound of wool in a shipment. In the latter case, however, the usual practice is to take some samples, examine them, and assume that the entire shipment is like the samples. Similar procedure is followed in mental testing. The measurement of ability to concentrate for a few minutes on a test blank is assumed to be a reliable sample of the individual's ability to concentrate for a prolonged period at his daily task; the average speed with which one operates a telegraph key when a light flashes is taken as indicative of his quickness of reaction when driving through traffic; or a sample of memory ability as manifest in a brief test is assumed to be typical of the person's

<sup>1</sup> In psychological terminology the "subject" denotes the person on whom the experiment is being performed or who is taking the test.

memory for the details of his business. Another feature that characterizes most of the better tests is the quantitative nature of the results. The person's score is not expressed as "good," "average," etc., but as a certain number of points. A mental test then may be roughly described as a scientific device for measuring quantitatively a typical sample of mental or motor performance in order to predict what an individual will do under certain circumstances.

Two things are rather essential in the psychologist's preparation for employment work. He must be familiar with the technique of test administration, that is, he must know how to use his tools just as a carpenter must know how to manipulate a saw. Mental test technique is the subject of the next chapter. But the psychologist also needs to know what tool to use on a particular occasion. It may be as ineffective for him to use a test of reasoning in order to predict ability at operating a hand-feed dial machine as for a carpenter to cut off the projecting end of a 2×4 beam with a hammer. And just as we should consider a man a very poor carpenter for attempting to smooth a plank with a chisel, in ignorance of the fact that there were available planes which would do a much better job, so a psychologist lays himself open to a similar charge of inexcusable ignorance if he uses an archaic and unreliable mental test when better ones are available. Considerable numbers of tests have been devised and perfected in recent years and most of them are accessible in the scientific literature. A person entering upon a project in employment research may find that some of these suit his purpose admirably, or at least that they will afford valuable suggestions for developing his own tests.

An employment psychologist thus needs a familiarity with a considerable range of mental tests that have been developed in various connections. Some of these tests will be illustrated in the present chapter. This discussion, however, will not constitute a miniature manual of tests. None of the examples will comprise a complete test, but merely enough items to illustrate its nature. Neither will standards nor the relation of tests to occupations be given in the present connection. It is usually necessary, anyway, to recalibrate the tests in the particular employment situation that is under consideration. The effort will be merely to give the reader some

notion of the types of mental tests that are available for the employment psychologist.<sup>1</sup>

#### CLASSIFICATION

**Capacity vs. proficiency.** The distinction between measures of capacity and measures of proficiency has already been made. The former deal primarily with innate or hereditary factors, while the latter are concerned essentially with acquired abilities. In the employment situation the former are used in predicting ultimate success in some kind of industrial performance in which the applicant, at the time of testing, has had no experience, while the latter are used to measure trade skill — i.e., the occupational ability which the person possesses at the time of application rather than some level of vocational proficiency that he will ultimately attain.<sup>2</sup> The tests for proficiency — i.e., trade tests — will be discussed separately in a subsequent chapter because those for different trades differ markedly from one another. Some of the tests for innate capacity to be described in the present chapter are more ubiquitous and may well be used for vocational prediction in many different lines.

**General vs. special capacity.** Tests of innate capacity may be further classified into those involving special capacity and those involving general capacity. We encounter situations in which a workman needs some rather special capacity, such as good attention or memory, quick reaction time, or accurate judgment of distances, in order to achieve success in his work. There are other situations in which there seems no outstanding special capacity like this necessary, but the person merely needs to be up to a certain general intellectual level, to be rather generally alert and able to adapt himself to circumstances — the thing that is often called intelligence.

<sup>1</sup> No consistent effort will be made to indicate the originator of a particular kind of test. In most cases this would be impossible because the tests have been repeatedly modified since their origin and they have appeared in scientific literature to such an extent that they are practically common property.

<sup>2</sup> A common type of proficiency test, which plays, however, little rôle in industry, is the standard educational test which measures proficiency in school subjects such as history or geography.

## TESTS OF SPECIAL CAPACITY

Practical justification of terminology. It is rather common practice in dealing with special tests to speak of them as tests of attention, tests of memory, and the like. This does not mean to imply, however, that the mind can be divided into clean-cut categories of this sort nor that a test can be devised that samples one of these categories to the exclusion of all others. While such terminology may be undesirable for theoretical purposes, it is justifiable for the practical employment situation. The practical man has a better notion of what the investigator is driving at if he speaks of measuring the clerk's attention or speed of decision than if he discusses test A and test B. The employment psychologist is not evolving a theory of attention or of judgment and it is not necessary for him even to define these terms which he uses. He is simply selecting certain tests that measure some aspect of mental performance, and the crucial point is whether these tests will enable him to predict an applicant's ultimate vocational success. He can call the particular tests used anything he wishes without affecting their utility, but he usually gives them a name that has a definite connotation to most persons and which probably has *some* relation to the thing actually measured by the test.

It is probably impossible to devise a test which measures one mental aspect to the exclusion of all others. Calling a test a memory test does not imply that it measures memory exclusively. If a person hears a list of words and then tries to reproduce the list, his efficiency will depend not only on his memory but on the extent to which he pays attention to the original reading. But this test will obviously involve memory to a greater extent than will a test in which the subject crosses out every letter A on a printed page. Thus, if a job rather patently necessitates good memory for its successful performance, it is desirable to try out some "memory test" which will probably measure that ability better than will a "decision test." In the following discussion, then, of tests for special capacity under different class headings, it must be remembered that these headings are used merely for practical convenience, and that tests do not measure exclusively the thing indicated, but simply emphasize it more than other things. After all, the real

problem is the correlation of test score with vocational ability regardless of what is actually measured by the test or what the test is called.

In the following pages a number of the categories of special capacity rather extensively used by employment psychologists are given with one or more examples for each category. The list of categories is not intended as exhaustive and only enough examples are given to illustrate the variety of tests in use. Where specific time limits or the quantity of items constituting the test are mentioned, this is not intended as an arbitrary suggestion, but is made merely for illustrative purposes. A person working in this field will usually go to original sources for his test material or else devise his own along lines suggested by the work of others.

#### MOTOR CONTROL

Many industrial operations involve coordination between eye and hand. The worker has to make certain motions, controlling them by what he sees. There are two aspects of motor control that are of significance, preventing motion — i.e., steadiness — and making motions accurately and rapidly. Lack of the former would seriously handicap, for instance, a jeweler assembling a small watch, while ineffectiveness in the latter on the part of a telephone operator inserting plugs in jacks would lead to still more wrong numbers.

**Example 1.** The conventional steadiness test makes use of a metal plate pierced with round holes ranging in diameter from one half to seven sixty-fourths of an inch. A needle or a piece of small wire is mounted on the end of a wooden rod so that the subject can hold it in the manner of a pencil and insert the point in the holes in the plate. The needle and plate are connected in series with a battery and an electric counter or other recording mechanism so that when the needle touches the plate the circuit is closed and the counter registers. The subject tries to hold the point of the needle in the hole for a prescribed number of seconds without touching the edge, beginning with the largest hole and working toward the smallest. A subject who can negotiate a smaller hole than another subject is obviously the more proficient in this particular test and the size of the hole may constitute his score. Other methods of scoring

are, of course, possible. For instance, a vibrating spring may be arranged to interrupt the battery circuit so that whenever the needle is in contact with the plate an electric counter is recording five times a second. The essential point is that a subject who is capable of preventing undue motion of his hand will make a better score whatever the method of administration.

**Example 2.** For indicating speed and accuracy of coördination a board is provided with three metal discs about one half inch in diameter mounted at the corners of an equilateral triangle four inches on a side. The subject holds a stylus (metal-pointed handle similar to a large pencil) with which he taps the discs in succession going around the triangle repeatedly in one direction. Each tap records electrically on a counter. The examiner can easily note the number of circuits of the triangle and measure the time with a stop-watch while the counter records the actual number of electrical contacts made. The subject may go as rapidly as he can and be scored according to his attempts and correct responses or he may be required to keep time with a metronome while the number of errors is noted.

**Example 3.** A metal plate replaces the record on a phonograph. There is a disc of insulating material about the size of a quarter set in this plate near the margin and flush with the surface. The subject holds a handle which terminates in a small brass point. The handle is hinged so that gravity keeps the point in contact with the metal plate when the hand is held over the apparatus. When the point is in contact with the plate an electrical circuit is closed. As the plate revolves, the subject tries to keep this point on the insulated portion so as not to make contact with the metal. During any time that he is making such contact an electric counter is recording ten times a second. At a given speed of rotation and for a given length of time, the subject with better coördination makes a smaller score on the electric counter.

The foregoing tests of motor control are necessarily individual tests, i.e., they must be given to one subject at a time. There have been efforts to devise methods of measuring motor control with group tests in which a number of persons are tested simultaneously. One such test will be cited.

## Example 4.



etc.

A metronome is set for some constant rate such as two beats a second. The subject starts with his pencil at the circle numbered 1. The signal to begin is given during some beat and on the next one the subject draws a line terminating as nearly as possible in circle 2; on the next beat he draws to circle 3, etc. The circles run in this fashion the full length of the page — perhaps thirty of them. The subject is stopped after thirty beats, and if he has not kept time with the metronome he will not have marked all the circles and can be penalized accordingly. The extent to which the end of each line he has drawn deviates from the circle in which it is supposed to terminate is inversely proportional to the subject's accuracy of coördination. Instead of measuring this deviation from each circle with a ruler, an arbitrary limit may be selected such as three thirty-

seconds of an inch and any line terminating within three thirty-seconds of the circle be considered correct. Scoring may be facilitated by a stencil with thirty circles of this radius drawn in the proper positions, and, by placing this stencil over the test blank, lines terminating outside the arbitrary limits may easily be noted. The test may be given with the blank in such a position that the zigzag lines will be made in a horizontal or vertical direction. Several trials at a test of this sort should presumably be given and the total number of correct responses or errors noted. Unlike most coordination tests this one may be administered by the group method.

#### SENSORY CAPACITY

**Example 5.** *Visual acuity.* The ordinary chart used by opticians with groups of letters of varying size gives a rough measure of visual acuity. The smallest letters that the subject can read at a distance of twenty feet indicates his acuity. For finer measurements letters of constant size are placed at such a distance that they are illegible and are moved toward the subject until he can read the letters. This maximum legible distance is then noted. Frequently a single symbol, such as the letter **E**, is used and turned with the opening pointing in various directions and the maximum distance found where the subject can correctly state the direction. These latter methods have the further advantage that they are less subject to coaching or other preliminary preparation. A very near-sighted scientist going into personnel work passed the army examination by purchasing all the different optical charts in the city and memorizing them so that when he saw the large letter at the top (which he could barely read at twenty feet) he could recite the remaining invisible letters. In his particular job the use of glasses was little handicap, but in an employment office there might be situations where such a procedure would be disastrous.

**Example 6.** *Color blindness* is roughly determined by the use of a standard series of small skeins of yarn dyed various colors. A certain green skein is given to the subject and he is told to select all those of the same color. Inasmuch as the colors are of numerous tints and shades there is considerable opportunity for a person who



cannot distinguish red from green to manifest this fact. More elaborate methods of determining color blindness are, of course, available if their use in employment procedure seems warranted. Color blindness may be present where the victim himself does not suspect it and where even his acquaintances are unaware of his defect. In some instances the first intimation has come when the man has married and a striking change in neckties has resulted. In certain jobs, of course, inability to distinguish readily between red and green might be fatal.

**Example 7. Auditory acuity.** A small steel ball is placed at a constant distance from the subject's ear and dropped a variable distance upon a metal plate. The ball is held by miniature pliers and released by slight pressure. The height of the pliers above the plate can be varied by turning a screw and the distance read on a scale. The minimum distance the ball can drop and still be audible is the differential score.

**Example 8. Pitch discrimination.** The Seashore series of phonograph records for measuring musical talent (528) includes one for pitch discrimination. Pairs of tones are presented differing slightly in pitch, and the subject in each instance determines whether the first or second tone is of higher pitch. He checks his results on a standard blank and the results can be compared with the known difference of pitch to determine his discriminative ability.

**Example 9. Kinæsthetic (muscle sense) discrimination.** The subject presses down on a spring scale, such as a postal balance, until the indicator as seen by the examiner reaches a certain point. The subject then is required to reproduce this pressure by remembering the kinæsthetic sensation, and his error is recorded. Technically superior devices of this sort, in which the subject produces variable tension on a spring by turning a crank with a scale that can be read more finely, have been devised.

#### ATTENTION

**Example 10.** A frequently significant aspect of attention is its range or span; i.e., the number of impressions to which a subject can attend simultaneously. Some mill operatives have, for instance, to watch several machines at a time. Some form of short-

exposure apparatus is necessary for such a test. One type comprises a shutter containing a slit which is pulled by a spring across the exposure field exactly like the focal-plane shutter in a graflex camera. The card containing the material that is to be presented is placed in a rack behind this shutter and as the slit moves across the field the material is exposed for a fraction of a second. The speed of exposure can be regulated by the width of the slit and the tension on the spring.

X P H  
 W K R M  
 L Z F D J  
 B S N Y F Q  
 G V L T H N R

With such apparatus it is possible to expose a series of disconnected letters like the above, one line at a time and thus to determine how many the subject can read in this fraction of a second. He is given 4 disconnected letters, then 5 letters, then 6 letters, etc., until he reaches his limit and the procedure is repeated several times. The maximum number that can be perceived during this brief exposure is termed the span of attention. Disconnected numbers or words or even nonsense figures or pictures may be used.

**Example 11.** More frequently we are not so much interested in the span of attention as in the ability to sustain attention for a longer period. In a clerical job we are perhaps concerned with ability to work continuously at a high degree of attention. A portion of a test for measuring this aspect follows with the first few items correctly marked:

1472859196325376935473984612348557216473189382576319272455245648 etc.

The subject underlines the pairs of adjacent numbers whose sum is 10. An actual test blank comprises 15 lines of this sort with 10 pairs to be marked in each line. The number marked within a definite time limit gives some indication of ability to work at a high degree of concentration.

**Example 12.** A more complicated test similar to the foregoing involves a page of numbers printed similarly, but the subject crosses 2 and rings 3 until he comes to a 7, then reverses the process, crossing 3 and ringing 2, till he comes to another 7, then reverses again, etc.

**Example 13.** A presumably somewhat different aspect of attention is involved in the following test in which the subject finds the consecutive numbers in order, i.e., finds 11, then finds 12, etc.

26	52	39	24	53	37	16
14	33	18	47	12	21	56
49	44	59	29	55	31	42
35	20	11	50	15	46	27
58	41	28	38	57	34	48
30	23	54	45	19	13	40
17	32	36	25	51	43	22

As a check on whether the subject actually finds the numbers consecutively rather than skipping around, he writes "A" after 11, "B" after 12, "C" after 13, etc. If he marks the numbers in any other than the correct order, he is apt to become confused in attaching the proper letter to the numbers and may be detected and penalized accordingly.

#### LEARNING

**Example 14.** One of the conventional tests for briefly determining ability to learn or to form a new set of associations involves the substitution of symbols from a code for a series of numbers.

1	2	3	4	5	6	7	8	9														
&	?	%	)	\$	f	(	"	/														
1	3	7	9	2	4	6	8	2	5	9	6	3	8	1	7	4	8	1	5	2	9	4
&	%	(	/	?																		
3	1	7	5	2	9	6	4	8	5	2	6	9	8	3	7	1	4	1	8	2	5	4

etc.

The subject writes under each number the corresponding symbol from the code at the top of the page as shown by the first few items. Enough blank numbers are provided to occupy the subject for the desired length of time. At the outset, of course, reference is made to the code for every number, but a subject who learns readily will soon remember some of the symbols without reference to the code and hence will work more rapidly and make a higher score. Various codes are, of course, possible, and if greater complication is desired the entire alphabet may be used and a symbol substituted for each letter.

**Example 15.** Another type of learning involves finding a pathway through a maze or labyrinth. Such a test is often given by a printed plan or diagram of the maze and the subject traces with a pencil the correct pathway. The following maze is similar, but may be set up on a typewriter or with ordinary printing equipment without the preparation of a plate.

```

A C C C X X C X C C C C C C C C C C
C X X X C X C X X X X X C X X X X C
C C C C C C C X C C C C C C C C X C
C X X X C X X X C X X X X X X X X X C
C X X X C C C C C X X X X X X X X X C
C X X X X X X X X X X C C C C C C C C
C C C X X X C X X X C X X X X X X X
C X X C C C C C C C C C X C C C C C C
C X X C X X X X X X X X X X X C X X X X C
X X X C X X C C C C C C C C X X X X C
C C C C X X C X X X X X X C X X X X C
C X X X X X C X C C C X C X X X X C
C X X X X X C X C C C X C X X X X C
C X X C C C C X C X C X X X X X X C
C X X C X X X X C X C C C C C C C C
C X X C X X C X C X X X X X X X X X
C C C C C C C X C X X X X C C C C C
C X X C X X X X C X X X X C X X X C
C X X C X X X X C C C C C C X X X Z
    
```

The subject starts at *A* and traces a continuous line to *Z* keeping on the letter *C* and always moving the pencil sideways or up and down; i.e., never moving diagonally. If tests of the maze type are repeated, the improvement gives some indication of learning ability.

## ASSOCIATION

**Example 16.** While free association tests are occasionally used in which the subject is given a stimulus word and then speaks or writes words as rapidly as he can think of them, it is usually more valuable to control the association process in some way, such as having the subject give or select synonyms or opposites of certain words.

1. RETURN is the opposite of ADVANCE; SURROUND; RESOLVE; GO
2. ACTIVE is the opposite of PERSON; PASSIVE; NEUTRAL; DESPONDENT
3. CONVEY is the same as CONDUCT; TRANSPORT; LIFT; GUIDE
4. OPERATE is the same as REFINE; DISTILL; SURGEON; MANAGE
5. CHARITABLE is the opposite of UNTRUE; ACT; MISERLY; UNFRIENDLY

etc.

The subject underlines that one of the four alternatives that correctly finishes the sentence. The first two lines are correctly marked.

**Example 17.** Another widely used test that may perhaps be classed here involves analogies.

1. Gun: shoots:: knife: RUN; CUTS; HAT; BIRD
2. Handle: hammer:: knob: KEY; ROOM; SHUT; DOOR
3. Camp: safe:: battle: WIN; DANGEROUS; FIELD; FIGHT
4. Egg: bird:: seed: GROW; CRACK; PLANT; GERMINATE
5. Cloud-burst: shower:: gale: BATH; BREEZE; DESTROY; WEST

etc.

The subject underlines that one of the four alternatives that is related to the third word in the line as the second word is related to the first.

## REACTION TIME

The measurement of reaction time was one of the early psycho-

logical experiments. Attention was first called to it by discrepancies of two astronomical observers recording the passage of a star across the meridian. One of them was consistently one second behind the other. This led to the discovery that people differ in the length of time required for an impulse to get in the eye or ear and out to a muscle. Such differences are now measured very accurately.

**Example 18.** To determine simple auditory reaction time the subject is given a warning signal "ready" and thereupon presses a telegraph key or similar device and holds it down. A few seconds later the experimenter throws a switch which causes something similar to a telegraph sounder to give a fairly loud click. The subject releases his key the instant he hears the click and the time between the click and the release is measured — usually to thousandths of a second. This type of measurement involves somewhat complicated technique and is obviously possible only in an individual test. There are various chronoscopes or time-measuring devices of this sort in use in psychological laboratories. One of the best of these comprises a synchronous motor controlled by an electrically driven tuning-fork so that it runs at very constant speed. A friction clutch operates electro-magnetically so that a disc may be held firmly against an extension from the shaft of the motor. This disc runs another shaft which carries a large hand. This latter revolves in front of a dial graduated into 100 units. The apparatus is so wired that when the stimulus or click occurs, it automatically sends current through the magnet which pulls the disc against the motor shaft. The hand then revolves ten times a second. When the subject releases his key this circuit is broken and another made which throws out the clutch and stops the hand with no further rotation. By reading the position of the hand on the dial before and after the reaction, it is possible to determine how many thousandths of a second were required for the subject to respond to the auditory stimulus. There are, of course, cruder devices for measuring reaction time. A less elaborate but fairly satisfactory apparatus consists of two small pendulums so adjusted that if they start to swing simultaneously the more rapid will gain a fiftieth of a second each swing. It can be so arranged that the two are held at

one side, the slower released by the stimulus and the more rapid by the reaction. By counting the number of swings until they synchronize, the reaction time can be determined to fiftieths of a second. Considerable time is consumed, of course, in counting the swings, and there are more chances of error in measurements made with this technique because of the difficulty of accurate adjustment of the pendulums.

**Example 19.** Simple visual reaction time may be measured in similar fashion. The stimulus may be the appearance or disappearance of a light or of a shadow or the motion of some small object. In any instance the stimulus automatically starts the chronoscope and the response of the subject stops it.

**Example 20.** Choice reaction time differs from the foregoing in that two or more alternative stimuli are given and the subject has alternative responses which he makes according to which stimulus is presented. For instance, there may be two telegraph sounders or similar devices, one at the subject's right and one at his left. He has two telegraph keys, and if he hears the right sounder he operates the right key, and if he hears the left sounder he operates the left key. Similarly there may be an arrangement whereby two shadows of a rod are thrown on an illuminated screen. If the right shadow disappears, the subject operates the right key, and if the left shadow disappears, the subject operates the left key. With more stimuli and more keys it is possible to complicate the choice to any desired extent.

#### SPACE PERCEPTION

**Example 21.** If a picture of a rhomboid-shaped card with two holes punched in it near adjacent corners is shown in two different positions, it is rather difficult to tell whether you are looking at the same or different sides of the card. On the test blank many pairs of this sort are provided, and the subject checks each pair to indicate whether it represents the same or different sides of the card. A somewhat similar test involves pictures of a human hand in various unusual positions, the subject in each case indicating whether it is the right or the left hand.

**Example 22.** A line several inches long and a much shorter line

appear close together. The subject judges without measuring how many times the shorter line is contained in the longer. A page of such pairs is provided. The subject may be required to write the number of times he thinks the smaller is contained in the larger or to select the correct number from several alternatives.

MEMORY

Example 23. *Memory span.*

8 5 7 3  
 9 4 2 1 5  
 7 3 2 6 4 9  
 2 9 5 3 8 7 1  
 5 9 7 4 8 6 1 3  
 8 3 1 5 7 4 9 2 6

The first row of numbers is read aloud at the rate of one digit per second. The subject listens during the reading and then immediately writes the numbers from memory. This procedure is repeated with the next row (five digits), then with six digits, seven digits, etc. The subject's score is the maximum number of digits that he can reproduce after one presentation. Several lists like the above are, of course, used. The method may be varied by having the numbers printed, each row on a separate card, and showing them to the subject for a length of time sufficient to allow about one second for reading each digit. This involves visual rather than auditory memory span.

Example 24.

book	shelf
garden	spade
letter	stamp
watch	time
rain	umbrella

etc.

The examiner reads the pairs of words rhythmically. A metronome sounds one beat a second and the examiner reads "book" on



the first beat, "shelf" on the second beat, pauses on the third beat, reads "garden" on the fourth, "spade" on the fifth, and "letter" on the seventh, etc. This serves to group together the two words of each pair and the subject is required to remember these two together. As soon as the list of perhaps twenty pairs has been read, the examiner gives the first word of each pair and the subject writes down the one that went with it. For instance, the examiner says "book" and the subject writes "shelf" if he can recall it; the examiner says "garden" and the subject writes "spade." The subject may have a blank with simply the numbers 1, 2, 3 . . . and write his answers after the proper number. The examiner, in reading the first word of the pair, allows five seconds for the response, whereupon he immediately passes to the next word. Or the subject may be provided with a blank containing all the first words of each pair and be given a certain time to write down all the second words he can recall. He may even have a blank of this form:

book.....page; shelf; title; case  
 garden.....flower; lawn; spade; plant

and be required to check that one of the four alternatives that was previously presented with the word at the first of the line. The same sort of experiment may be conducted visually, showing the pairs in succession at a small window in specially constructed apparatus. They may be typed on adding-machine ribbon, each pair on a separate line, and fed along in guides behind a slit in the apparatus by pulling the ribbon. This usually necessitates individual examining, but it is possible to place this apparatus in the exposure field of some projecting lantern such as a Balopticon and throw the words on a screen at the front of the room in which a group of subjects is sitting.

**Example 25.** A paragraph, such as a newspaper account or a description of some scene, is read aloud, and the subjects then either reproduce it in their own words or are asked specific questions requiring unequivocal answers to test their memory for details of the selection. In the former case the selection is divided into "ideas" and the number of these reproduced is counted. In

the latter case the number of questions answered correctly furnishes the index of memory for the selection.

REASONING

**Example 26.** A series of arguments like the following is given and the subject marks an item *X* if the conclusion is true and marks it *O* if the conclusion is false.

- ..X..1. John's birthday is after Harry's and Harry's birthday is after Tom's. Therefore Tom's birthday is before John's.
- ..O..2. William has a brother George who has a son, Henry. Therefore Henry is William's uncle.
- .....3. Silver is heavier than iron. Copper is lighter than silver. Therefore copper is heavier than iron.
- .....4. Jones owes Smith one hundred dollars. Brown owes Jones one hundred dollars. The two debts will be settled if Smith pays one hundred dollars to Brown.
- .....5. All members of the Country Club are members of the Polo Club. Smith is not a member of the Polo Club. Therefore he is not a member of the Country Club.

**Example 27.**

A O U U A	A A O U A U	A A U O U A
1. U O A U A	2. O U U A A A	3. U A A O
A O A A	A A A A O U	U U U O A U
U A U U U A O A	A U U A O A A A U	A O U U U
4. U U A U U A O U	5. A O U	6. U U A O A
A U A O U U A	A U O A A	U U U U A O
etc.		

The letter *O* in each line bears a certain relation to the rest of the line. The same relation holds for all three lines of the given problem. For instance in the first problem the *O* is "second from the left" in all three lines. In problem 2 the letter *O* occurs "before the first *U*" in all three lines. In problem 3 the answer is "fourth from the left"; in problem 4 it is "after the second *A*." The subject writes these phrases under each problem.

## SPEED OF DECISION

## Example 28.

OOAEUAUAA	OEAOUEOOUA	EAOEUEOAOE	UUEOUAOEOE	EUGAEEOUAE
OUAAAEOEAO	UOUAQQAEUO	EUUEAAOEOA	OAUAEQUEUA	QPAKEUEAE
AUAAEAQOAE	OUOEEQAOAO	OEUAEEEAUU	UOEAEUQUU	EAEEOUUEU
UUAOEAAAUU	EOEAQOAUOU	EAOEAOEOUE	QAUAAAQOEU	AEQEEEOEOE
AUAEEAEEA	OEQOEQUAQQ	EAEUQOEEU	EOAUQOEUU	EAEQOEUAE
[A]	[O]	[E]	[ ]	[ ]
AEOUQAEQO	UAUEUAUEUA	AOEUGAEQOE	EAUEQAUQOU	AFAQEOAUA
QOEQOQOOU	AQUAQUAQOA	QOQOQUQUU	UEQOQOQOE	QAUAAUAEU
AQOEAQOQA	QUUQUQUUEU	AUQEAQOQU	EUEAQUEAQE	AFAEAQAAA
QUAQEOQOE	UEAUAEUQOE	QAEQOAEQAO	EAUEAUQOEA	QAEUAQOEAU
QUQOEAQOU	EUQEUUEUU	EQAUQAEQOU	AEUAEUQOAE	QAEUAQOAU
[ ]	[ ]	[ ]	[ ]	[ ]

etc.

The subject is allowed five seconds to glance at each square and determine which letter predominates. The result of this quick decision is written in the brackets below the square. A typical blank comprises 48 squares of this sort. The examiner gives the signal "Begin" and in five seconds says "Mark." Thereupon the subject immediately writes his judgment in the first bracket and looks at the next square. Five seconds later the examiner again says "Mark" and the subject immediately writes under the second square and turns to the third. After the examiner has said "Mark" 48 times the subject is prevented from writing further, so that if he did not keep up with the examiner there are some unmarked squares to reveal that fact.

## INGENUITY

## Example 29.

Animals and birds		Fruit and vegetables	
eehps	beelt	aelpp	inprt
ekmnoy	binor	saabnn	acnp
abers	eginop	aegpr	amoot
ehnot	kknsu	alntuw	abens
aekns	aeelsw	elmno	acorrt

etc.

The letters of each word are arranged alphabetically rather than in the normal order. Those in the first group are names of animals or birds and those in the second group fruits or vegetables. The subject determines what word the letters would make if put in the correct order and writes it after the corresponding letters. He is given a short time limit for each group of words and in that time skips around and gets as many of that group as possible. Other categories such as proper names, furniture and cities may be used.

**Example 30.**

- |                 |              |              |              |
|-----------------|--------------|--------------|--------------|
| 1. <u>s</u> pot | mind         | long         |              |
| 2. <u>b</u> all | <u>m</u> eat | <u>s</u> and | <u>f</u> our |
| 3. sift         | play         | army         |              |
| 4. twig         | hope         | fill         | flag         |
| 5. hand         | note         | grab         |              |
- etc.

In each problem, if one letter is taken from the first word, one letter from the second, and another letter from the third, and they are put together in that order, they will form the name of an animal. In the first line the three underlined letters spell *pig*. In the second line the answer is *bear*.

**Example 31.**

Ann t... the poker and began breaking the big l... of coal in the g... as she said this. Little spirals of greenish yellow s... escaped from the cracks made by the p... then jetted into f... She was so s... for this woman before her that she l..... doggedly at a lump of coal a.. the w.... that she was speaking.

The subject fills in the blanks in the text. The test may be varied by having the number of missing letters indicated by the dots or other symbols, as in the present instance, or by giving no clue as to the length of the word. The initial letter may or may not be given.

## ABILITY TO FOLLOW DIRECTIONS

## Example 32.

- If the word contains the letters E, A, and R mark it 1.  
 If the word contains the letter E but not A and R mark it 2.  
 If the word contains the letter A but not E and R mark it 3.  
 If the word contains the letter R but not E and A mark it 4.

years.....reason.....taint.....addition.....  
 height.....action.....beguile.....island.....  
 alfalfa.....bright.....bureau.....office.....  
 right.....print.....rough.....when.....  
 verbal.....rocky.....lurid.....forbear.....

etc.

The test can be made more complicated by using other combinations such as "E and A but not R"; "A and R but not E." In general the more combinations involved, the more difficult will it be to follow the directions.

## Example 33.

Make a cross here ... and a circle here ... and cross out the second and third letters of this word — PECULIAR. If you think there was a war in 1917 put in a number to complete the sentence: "A horse has ... feet." If Tuesday comes after Monday make two crosses here ... but if not make a circle here.... If it snows hardest in summer make a cross here... but if not pass on to the next question and tell what you wear on your hands in cold weather. .... Draw a line between the names of these two boys George Henry and write "no" if 2 times 3 are 6. Notice these five letters A B C D E. Draw a line from A to D that will pass above B and below C. Notice these numbers — 3, 5. If a rock is heavier than a feather write the larger number here..... but if not write the smaller number here..... Give a wrong answer to the question "How many days in a week?" ..... If sand is good to eat write "no" here ... but if it is not, write "yes" here.... If fishes live in the water make a triangle here .... and a square here.... Cross out every letter E in the words between triangle and the square which you just drew.

etc.

The foregoing illustrations have dealt with factors where a fairly satisfactory objective measurement has been possible. Psychologists and particularly the industrial psychologists have come to realize recently that we do not have here the whole story. Fre-

quently a person has the requisite ability for a job, but fails to succeed because of some emotional difficulty, because of dishonesty, because of some temperamental factor, or because of lack of interest. It is a question not merely of what the man *can* do, but of what he *will* do. These discrepancies are often rather serious in the practical situation, for they not only render prediction less reliable, but often unjustly reflect discredit upon the tests. The ability measured by a test may be related to vocational aptitude, and if the employee does not use that ability it is no fault of the test, although the test often gets the blame. This emphasizes, however, the desirability of obtaining further information about these less tangible factors. Only recently have serious efforts been made to measure them and a few typical attempts along these lines are illustrated below (the question of interests is left for a separate chapter). It must be emphasized, however, that these attempts are still largely in the experimental stage and should not be regarded as at present validated, but should rather be considered as suggestive of the lines along which work in the near future will progress.

#### EMOTION

**Example 34.** The subject has around his chest a pneumograph — a large soft rubber tube supported by a spiral spring inside — connected through a smaller tube to a small brass cup or “tambour” covered with a thin sheet of rubber. As the chest expands or contracts, the air pressure in the pneumograph and the rest of the system changes so that the sheet of rubber over the cup moves up and down. As it moves it pushes a light lever. This latter presses lightly against the surface of a rotating drum covered with smoked paper. The result is that a wave-like curve is traced on the drum recording the subject’s inhalation and exhalation. The subject also holds in his hand a long handle carrying another little cup covered with sheet rubber and with a top-heavy piece of brass glued on the rubber. If the subject’s hand trembles, the piece of brass moves sufficiently to displace the rubber and change the pressure in another similar tambour with a writing lever. Thus, if the subject holds his hand still, the lever traces a comparatively straight line on the smoked drum, while if his hand moves, the line becomes

irregular. The subject with these, and possibly other attachments, sits quietly and then a revolver is fired a short distance behind him. The subject knows it is to be fired during the experiment, but does not know just when. The smoked record usually shows a sudden vigorous motion of the hand and a quick deep breath. However, there are individual differences in the speed of recovery from the shock. The breathing of some subjects after the first involuntary gasp steadies down to normal depth and rhythm almost at once, while other subjects take half a minute or longer to recover. Similarly, with some the record for motion of the hand traces a straight line with only the instant's interruption, whereas with others there is a pronounced tremor for some time. The length of time before the hand curve shows as little irregularity as at the outset or before the breathing curve comes back to normal gives a somewhat quantitative indication of emotional stability.

**Example 35.** A different approach to the measurement of emotion involves the extent to which it distracts the subject from a task. The subject is given a standard amount of practice in some mental task, such as performing addition. For instance, if the problem is to start with a specified number, add 1, then add 2 to the sum, then 3 to that sum, such as 30, 31, 33, 36, 40, 45, 51, it can be determined how many practice series are necessary to bring the average person to approximately his maximum proficiency. The subject is given this amount of practice and then a normal series recorded. The subject then grasps a pair of electrodes while the examiner manipulates a bank of lamps. The subject is told that at various times during the test he will be given a shock. He is at the outset given a mild one and his attention called to the voltmeter which registers at the time say 75 volts. He then releases the electrodes and is shown how a further manipulation of the lamps will make the meter register 220. He then takes the electrodes and is given an initial number and proceeds to do his addition. If he is performing, for example, ten additions in a series, he may be given a mild shock at the end of the series so that the shock itself will not actually distract him from his task, but the fear of the shock is liable to do so. Consequently, the longer the time for series with electrodes relative to normal series without electrodes, the more sus-

ceptible the subject may be considered to the emotions involved.

#### HONESTY

The honesty of a prospective employee is frequently of prime importance, but its evaluation is withal a delicate matter. Such evaluation must be objective and without the subject's knowledge. It must not consist of "framing" him in such a way that he realizes what has occurred. He will tell his friends, and this will vitiate the subsequent use of the test as well as build up an attitude of resentment. It should, however, be possible, in what is ostensibly a conventional test of some mental or motor ability, to obtain indirectly information as to the subject's honesty without his ever realizing what has happened. Many persons may feel that this procedure is going too far and prying into sacred portions of the personality, but if an employer obtains information about a person's dishonesty from others who know him or from some objective manifestation of it, he uses that information. It would seem no more unethical to utilize such information when obtained objectively in the test.

**Example 36.** A series of circles are arranged on the blank with their centers all approximately on a large circle. These circles are of various sizes. The subject places his pencil at a designated point, shuts his eyes, and attempts to make a cross in each circle. He is given several trials scoring each trial himself before doing the next. By trying the test on persons who are actually known to have their eyes closed, it can be determined just what are the chances of hitting some of the smaller circles. If the subject does considerably better than this probable expectation, the presumption is that he "peeked." (108.) In another variation of this test there are six squares of different sizes one inside the other, thus affording a continuous pathway between each two squares. In each pathway the subject starts with his pencil at a designated point and with eyes closed traces around through the pathway to the starting point. With the shorter pathways a correct response is possible, while with the longer pathways it is perhaps impossible. This test can be evaluated in the same manner as the preceding one with the circles.



**Example 37.** The subject is given a set of preliminary questions such as:

1. Can you swim?
2. Can you skate on roller skates?
3. Can you drive a car?
4. Can you drive a boat?

etc.

In each instance he grades himself as to the matter in question, assigning himself a value of 3 if he can do it very well, 2 fairly well, 1 if he knows something about it, and 0 if he knows nothing about it. The preliminary set of questions is followed by a more crucial set which the subject answers in the same fashion:

1. Do you know the letters of the alphabet in their order?
2. Do you know how to write decimals?
3. Do you know what a fly wheel is for on a steam engine?
4. Do you know how a camera takes pictures?

Two or three weeks later, after the subject has presumably forgotten much of this preliminary test, he is given a test dealing with the information called for in the first case, such as:

1. What is the fourth letter after *M* in the alphabet?.....
2. Write four fifths as a decimal.....
3. Fly wheels are placed on steam engines in order to:

aid in stopping them ..... true...not true...  
 help them keep going ..... true...not true...  
 tell how fast the engine is going ..... true...not true...

A considerable number of items of this sort are used as a check on the previous statements of the subject in an effort to determine whether he falsely overstated his ability in the first test.

#### TEMPERAMENT

**Example 38.** The subject copies certain words, such as "United States," at normal speed. On another page he copies similar words as slowly as he possibly can while still keeping the pencil moving. Other words are to be written as quickly as possible, but keeping the writing legible. The subject then writes "United States" repeatedly while watching the examiner without looking at the paper and counting simultaneously the number of taps made by the examiner.

Other tests involve imitating a written copy, disguising one's own writing in repeated trials, and writing certain words in a very small space so that the letters must be somewhat crowded. Finally, on a list of traits, such as "careful — careless," the subject checks which one is generally the better, and later in the test goes through a similar list checking which of the two describes *him* better. Having given such a test it is possible to evaluate such things as ability to go very slowly and to inhibit or check the tendency toward greater speed; ability to release the usual inhibitions and operate very rapidly; tendency to change the character of performance under distraction; facility in adapting one's self to conditions, as when given inadequate room for normal performance; speed of deciding about personal traits relative to speed of deciding about such traits when the personal element of self-evaluation is not involved. (148.)

**Example 39.** Efforts have been made to get some measure of susceptibility to monotony as follows: The subject has an endless belt of small brass rings four inches apart connected by string and two small pegs on a board four inches apart. The belt runs under the table. The subject puts two adjacent rings over the pegs, then removes them and pulls the belt along to put the next pair over the pegs. This is done for a standard set of time intervals. A somewhat similar task is interspersed in which the distance between the rings is varied from four to twelve inches and the pegs are mounted on blocks so that they can slide to the right or left. These rings are painted two different colors and the distance between the pegs must be adjusted according to the distance between the rings. Moreover, if a red ring appears, the pegs must be placed in a position as far to the left as possible, while if a blue ring appears they must be adjusted toward the right. The subject likewise estimates the duration of the time intervals for which he works. The subject performs the task at whatever speed he wishes. However, the examiner keeps count of the number of strokes made every fifteen seconds to determine whether the subject slows down as the task progresses and whether this tendency is more marked with the uniform set of rings than with the variable set. The subject's time estimation may likewise give some measure of the monotony by indicating whether the time is "dragging."

The foregoing examples give some notion of the scope of mental tests of special capacity or of attitude. There has been no effort to illustrate all the possible varieties, for this would be out of the question in the present work. Effort has been made merely to show some of the possibilities and some of the kinds of test that have actually been used in one or another practical problem. A previous statement should be reiterated, that the class headings used are merely for practical convenience and do not presuppose a careful definition or analysis of the factor so classed, nor do they imply that the test measures that factor to the exclusion of other factors. The real problem is the correlation of test score with vocational ability regardless of what the test is called or what it actually measures.

#### TESTS OF INTELLIGENCE

We have thus far been illustrating, for the most part, tests of special capacity. There are many industrial situations where a person is being considered for a job that requires unusually good mental equipment along some particular line. He may not need all-round ability, but rather something quite specialized for the particular limited group of operations he is to perform. The tests thus far described are designed largely to meet this situation. On the other hand, as suggested earlier, there are situations in which there seems no outstanding special capacity of this sort necessary. The person needs to be generally alert, perhaps, and able to make moderate adjustments to the conditions of his job, but he does not need quick reaction time any more than he needs speed of association or ability to make quick judgments. We usually speak of such a person as being of a certain intellectual level or possessing a certain degree of intelligence.

**Nature of intelligence.** This is not the place for an elaborate discussion of the nature of intelligence, for, just as with the tests of special capacity, the crucial point is whether the particular tests facilitate vocational prediction, regardless of what in the last analysis they measure. Electricity could be measured and used for practical purposes before it was defined or its exact nature known. Similarly, the psychologist can measure intelligence for practical

purposes even though he is not certain what it is. Scientists' conceptions of intelligence appear to depend somewhat on the interest with which they approach the problem. A statistician is quite apt to think of it as a general factor which causes intercorrelations between miscellaneous mental tests; i.e., persons who make high scores in one kind of mental test often make somewhat similar scores in a good many different kinds of tests and *vice versa*. A person with biological interests may be inclined to conceive of it in terms of ability to adapt one's self to his environment and make the appropriate adjustments as the environment changes. A certain degree of flexibility in behavior seems demanded by human society. One with a physiological trend is inclined to think of intelligence with reference to its neural aspects — the degree of plasticity of the nervous system and facility in forming new connections and patterns therein. To the business man it connotes mental alertness, ability to follow instructions, to analyze a situation, to learn readily, and to "catch on" to new situations. This last conception comes nearest to that which would be adopted by an employment psychologist if it were necessary for him to commit himself as to the nature of intelligence. There appears to be some general capacity that gives one a better chance for survival in the economic struggle. One man can start out in any one of a dozen lines of work and be successful in any of them. Whatever line he attempts he gets a good start, he learns his duties readily, he adjusts himself to the situation, and does his work rapidly and accurately. Another man, though he may try many types of vocation, is practically doomed to failure in any of them. He is "dumb"; he does not "get the idea"; he cannot adjust himself; he is slow, and he often does the wrong thing or fails at least to do the right thing. This is the type that is floating around trying one job after another and losing it. As far as employment psychology is concerned, we may say that the first man has high intelligence and that the second man has low intelligence.

**Kinds of intelligence.** Most of the tests for intelligence thus far devised have been of the "abstract" type; i.e., the subject is asked certain questions or presented with certain problems that have an abstract ideational content. Moreover, the majority of these tests

are in verbal form; i.e., they necessitate the use of oral or written words. To be sure, much of the thinking and much of the achievement of the normal literate adult goes on in these terms. But there are individuals encountered in the employment office who will be at a distinct disadvantage in this type of test — foreigners who have difficulty in handling our language or persons who have had little or no experience in reading. Some such individuals may actually have good capacity of the sort usually called intelligence, but will fail to manifest it in the verbal form of test. For the benefit of such persons there have been developed various non-verbal or performance tests in which the subject does not have to deal with words.

There is another aspect of intelligence of which we must take some account. There are grounds for the suspicion that there is something which may be termed "mechanical intelligence." Some persons who do not manifest a high general capacity for handling abstract concepts may nevertheless have a distinct general superiority to their fellows when it is a question of manipulating concrete objects. Dealing with things that you can take in your hands and place in different positions and put together in different ways is somewhat different from dealing with words that are the opposite of each other or determining the relation between different pairs of words or successive numbers in a series. By this mechanical intelligence is not meant mere manual dexterity, nor ability to perform a single mechanical operation, but rather something of a more general character. Just as high intelligence of the abstract type enables a man to be successful in any one of many vocational pursuits that involve that kind of intelligence, so a person with high intelligence of the mechanical type will presumably be successful in any one of many vocations where he deals with concrete rather than with abstract things and manipulates objects other than a pencil. This notion of mechanical intelligence is not as firmly grounded as the other and less actual experimental work has been done in this field, but it is an aspect of which the employment psychologist should take account in certain practical situations.

There is still another type of intelligence with reference to which little has been done and that is "social intelligence." There may be instances where we should consider a person's general capacity not

for dealing with abstract concepts nor for handling concrete but inanimate things, but rather for dealing with social situations and reacting to other people. Tests of this type are still very much in the experimental stage, but if they are subsequently developed they may be of considerable practical significance for the types of vocation in which social contacts play a large rôle.

When we turn to illustrations of common tests for intelligence, especially of the abstract verbal type, we find that they involve little actually new material in addition to that which we have already encountered in tests for special capacity. The common practice seems to be to select a considerable variety of items of the sort described previously and lump them together into a single scale. If intelligence is a rather general characteristic, about the best procedure is to sample a fairly wide range of special capacities and assume that the combination of these will give some indication of general capacity or intelligence. In the remainder of the chapter it will be profitable first to illustrate the abstract verbal type of test — both that which necessitates individual examination and that which is adapted to group examination. Then the abstract performance or non-verbal tests will be likewise illustrated in the case of individual and group tests. Finally a few illustrations will be given of the mechanical types of intelligence test.

**Example 40.** The most widely used individual intelligence test is doubtless the Binet test as revised by Terman. (588.) It consists essentially of a series of questions for each age which the average child of that age can answer correctly. For instance, a series of questions are devised which the average five-year-old child can answer and a similar set which the average six-year-old can answer. These questions are then used to determine whether a given child's mental age is five or six. There is, of course, the possibility that a person will fail on a question for a certain age and compensate for it by passing a question for a higher age. The essential point is that the average child whose chronological age is five will test exactly five or have a mental age of five. Such questions are available up to the age of fourteen as well as an additional set of questions for "average adult" and "superior adult." Consequently, any person can be given these tests and his mental age determined. If he

passes only the questions that are attained by the average child of eleven years, he is said to have a mental age of eleven. By way of illustration the items for several different ages will be given. The nature of the item will be mentioned and where its administration is not obvious the words used in asking the question will be added.

The questions for three-year-olds are as follows:

1. Point to the nose, eyes, mouth, hair. "Show me your nose."
2. Name familiar objects — key, penny, knife, watch, pencil. "What is this?"
3. Look at standard pictures and enumerate at least three objects in one picture. "Tell me what you see in this picture."
4. Give sex. "Are you a little boy or a little girl?"
5. Give last name. "What is your name?"
6. Repeat a sentence of six or seven syllables. "I have a little dog"; "The dog runs after the cat"; "In summer the sun is hot."

The items for the nine-year-olds are as follows:

1. Give the date — day of week, day of month, and year.
2. Arrange in order five small cubes of uniform size weighing 3, 6, 9, 12 and 15 grams.
3. Solve problems in making change such as: "If I were to buy 4 cents' worth of candy and should give the storekeeper 10 cents how much money would I get back?"
4. Repeat four digits such as 6-5-2-8 backwards.
5. Make up a sentence containing three words such as "boy, ball, river."
6. Give at least three words to rhyme with a word given by the examiner such as "day" or "mill."

The following tests occur at the average adult level:

1. Define a certain number of words in a standard vocabulary list.
2. Interpret fables which are read by the examiner receiving certain credit for each one correctly interpreted and being required to make a certain total score.
3. Differentiate between such pairs of words as "laziness and idleness"; "evolution and revolution"; "poverty and misery."
4. Solve problems of this sort: "You see this box; it has two smaller boxes inside and each of the smaller boxes contains two tiny boxes; how many altogether counting the big one?"
5. Repeat six digits such as 4-7-1-9-5-2 backwards.
6. Learn a code for the letters of the alphabet based on a simple geometrical principle and then write a phrase in code without reference to the original copy.

**Example 41.** Among the verbal group intelligence tests that are most widely used is the "Army Alpha" test devised for military use during the war. It is based on the principle mentioned above that a considerable number of special capacity tests may be lumped together and give a fair indication of general capacity or intelligence. This test has served as the prototype for many other group intelligence tests that have been developed since. One such test may be cited. It is made up of seven kinds of items with at least thirty of each kind and in some instances many more. It includes the four kinds of items described below in connection with the discussion of test instructions (p. 123), namely, opposites, disarranged sentences, number completion, and analogies. It also comprises items of the following sort:

Get the answers to these problems. Write the answers on the dotted lines to the right.

1. If you walk four miles an hour for three hours, how far will you have walked? .....
2. If I had 50 per cent more money than I now have, I would then have \$84. How many dollars have I now? .....
3. A householder has food to last three people five weeks. How long will it last five people? .....

Look at the first word in the row. Underline one word in the same row at the right that the first is most often used to describe.

1. RED tree; rose; butter; milk; bottle
2. ABIDING fortitude; faith; hatred; anxiety; attitude
3. NAVIGABLE boat; sailor; navigation; stream; novice

Look at the statement at the left of the line. Underline one word in the same row which will finish the sentence and make the best sense.

1. People hear with the — eyes; ears; nose; mouth; hands
2. Chard is a — fish; lizard; vegetable; snake; fruit
3. The *Literary Digest* is published — monthly; daily; yearly; bi-monthly; weekly.

In this particular test complete directions and practice examples are provided on the first page of the blank. The items follow — each



page being devoted to a single kind. The subjects work continuously turning from page to page as directed by the examiner.

**Example 42.** Turning now to examples of the non-verbal or performance tests that involve individual examination, one of the common types is the "form board." This test appears in many varieties. For instance, a board is provided with holes of various shapes cut out of it — square, circle, cross, star, diamond. Blocks are provided of the proper shape to fit these holes. The subject's problem is to fit the blocks into the holes as quickly as possible and a record may be taken of the moves he makes and of the time. This particular board is perhaps too simple for ordinary industrial use, but the principle can be extended to include all degrees of difficulty. For instance, there may be a single rectangular hole and a number of rectangular and triangular blocks which if fitted together in the proper manner will fill the hole. Various other complicated patterns have been devised. The essential point is that the subject puts the blocks together in the proper fashion.

**Example 43.** Another common performance test involves "cube imitation." Four small cubes are mounted about two inches apart in a straight line on a wooden base. The examiner takes a fifth cube and saying, "watch carefully and do just what I do," taps the cubes in some predetermined order such as 1-2-3-4. The subject then imitates the moves made by the examiner. The series of moves may be complicated at will. For instance, if consecutive cubes are denoted by numbers 1 to 4, typical sequences would be:

1-3-2-4-2-3-1

1-4-2-3-1-3-2-4

**Example 44.** It is possible in some instances to administer a performance test as a group test. Such a test was devised in the army and administered by means of an examiner and a demonstrator. Demonstration materials are provided on a blackboard; the examiner in pantomime explains each kind of test item to the demonstrator, who then performs it on the blackboard in view of the subjects. The demonstrator acts the part of a high-grade moron. He apparently fails to understand so that the examiner must show him repeatedly. Then when he himself attempts the performance he

makes stupid mistakes, whereupon the examiner shouts "NO" and shows him all over again. Finally it begins to dawn on him, and he starts with hesitation, frequently watching the examiner for encouragement, and after finally succeeding registers great satisfaction. This "act" serves to center the attention of the subjects on the test material and the importance of doing it in the proper way. They are then told in pantomime and with a few standard phrases in the appropriate foreign languages to do the same thing on their blanks.

The first items consist of tracing with a pencil through the correct pathway in a series of simple mazes. These mazes or labyrinths are printed on the blank with a correct pathway by which the pencil can be drawn from start to finish without crossing any lines. There are various blind alleys into which the subject may go erroneously. A few illustrations appear on the blackboard. The examiner traces the correct pathway with his pointer and then the demonstrator does likewise with his crayon. The demonstrator makes a mistake going into a blind alley or crossing a line and is reproved by the examiner. After a few illustrations of this sort, the subjects usually understand their task and perform it on their blanks. The next items involve pictures of piles of cubes so arranged that some of the cubes are invisible, but their presence and the total number of cubes can be inferred from the arrangement. The examiner points to a sample on the blackboard asking, "How much?" and the demonstrator counts the cubes and writes the number in the proper place. The samples include cases in which some cubes are concealed. The subjects then study their problems and write the proper number under each. The next problems consist of completing a series of X and O symbols, such as OOXOOXOOX..... The blank has rows of rectangles in which the initial parts of the series are given and the subject fills in the remaining rectangles accordingly. Illustrative examples on the board are worked out by the demonstrator as the examiner points to them. The next items involve substituting symbols for digits as in Example 14 (*supra*, p. 70). It is explained in pantomime like the others. Further items in this performance group test involve checking certain pairs of numbers to show whether they

are the same or different, supplying missing parts of pictures such as the leg of a table or a finger of a hand, or indicating how a number of odd shapes can be placed together to fill a given square area. The items in this particular test are similar to those that have been used in other cases, but the novel feature consists in "putting it over" to illiterates. The tests are so adapted that reading is unnecessary. The subjects do not even need to know English in order to understand what they are supposed to do.

**Example 45.** One of the most widely used individual tests for mechanical aptitude or general mechanical intelligence consists of assembling a number of small appliances. A box is provided made up of ten compartments. The first contains the three parts of a small simple monkey wrench. The subject is required to put the wrench together by putting the head through the end of the handle and inserting a thumbscrew at the proper place. As soon as he finishes this compartment, he turns to the next in which there are six links of a light chain. These likewise must be assembled in correct fashion. The next compartment contains the parts of a spring paper clip. Other compartments involve a bicycle bell, a coin holder, a spring clothespin, a shut-off for a rubber hose, a push button, a simple lock, and a mouse trap. The subject assembles the items one after the other and usually has a time limit for the entire examination. The different items are scored according to a special scale. For instance, a perfect assembly of the wrench gives ten points; if the nut is in the wrong place, there are four to six points; if in addition the head is turned in the wrong direction, only one point is allowed. A similar scale is available for each assembled object. The total score gives some indication of general mechanical aptitude.

**Example 46.** The foregoing example is probably better adapted to men and boys than to women and girls. A test involving somewhat the same principle has been used to measure mechanical aptitude or intelligence in women. It consists of a series of envelopes each containing a sample and the materials necessary for making an object like that sample. The subject takes them in the order in which they are arranged and solves each problem. The first problem involves stringing twenty-four large colored beads to form

on the string a pattern of four yellow, four blue, four red, four yellow, four blue, and four red in that order. The next problem consists of putting a piece of tape through a strip of "insertion." In the third a card is provided with eight holes punched along the margin of a circle and an additional hole punched in the center. A string threaded on a needle has to be passed through the center and one of the outer holes back through the center and out to the next outside hole, back to the center again, etc. Other problems of constructing something like the sample include cross-stitches on a piece of checked gingham — the stitches coinciding with the squares of the fabric; assembling a simple key ring on a chain; making a chain of paper clips; sewing a piece of braid along the edge of a piece of cotton; assembling an address tag for a suitcase or grip; winding two strings around a card with notched edges in a certain pattern interlacing the strings at regular intervals; making a booklet with a piece of cardboard and stickers for hinges; and cutting out with scissors an irregular printed pattern bounded by lines one sixteenth of an inch apart keeping between the lines. This test may be scored in somewhat similar fashion to the preceding example.

**Example 47.** The two foregoing examples are essentially individual tests. If, however, duplicate sets of material are provided, they may be given to small groups simultaneously — the subjects solving the problems in order with a given time limit for the whole. To give them on a large scale as group tests involves a considerable outlay. Efforts have been made accordingly to measure somewhat the same factors by means of a printed blank that may be administered like the usual group test. One such test involves small pictures of a variety of mechanical objects. They are presented in groups of five pictures each. The groups are arranged in pairs and so constituted that each object in one group belongs with an object in the paired group. The objects illustrated in a typical group are as follows:

FIRST GROUP	PAIRED GROUP	
1. screwdriver	A. twist drill	1. . .C..
2. bit stalk	B. anvil	2. . .E..
3. tire pump	C. wood screw	3. . .D..
4. brace	D. tire	4. . . . . .
5. hammer	E. bit	5. . . . . .

The objects in the first group are numbered from 1 to 5 and those in the paired group are lettered from *A* to *E*. The subject must identify an object in the second group that goes with each object in the first. In the margin are the numbers 1, 2, 3, 4, 5. The subject writes after each number the letter for the corresponding object that belongs with it. In the above illustration after 1 he writes *C*, because the picture of the screwdriver and that of the wood screw belong together. After 2 he writes *E*, because the picture of the bit stalk and of the bit belong together. Another set of pictures involves in the first group a valve grinder, spark-plug wrench, a throttle, a set of coil points, and a hydrometer. The paired group contains an accelerator, a storage battery, a spark plug, an engine valve, and a spark coil. Other groups involve such things as locks, curtain rods, hinges, telephone construction, gauges, and parts of vehicles.

**Example 48.** One other type of intelligence test should be illustrated for the sake of completeness. This is the "omnibus" test. Instead of grouping the various kinds of test items so that a page of one type is completed before turning to the next, the different types are intermixed. The subject does a very few items of one sort, then a very few of another sort, etc. He may even do only one item of a given sort at a time, thus shifting very rapidly from one kind to another. The test described in Example 41 might be put in the omnibus form as follows:

**THICK** is the opposite of: **HEAVY; LARGE; THIN; SMALL**

dogs climb meat eat

22 24 26 28 29 30

bird: sings:: dog: **FIRE; BARKS; SNOW; FLAG**

If you walk four miles an hour for three hours how far will you walk? .....

**RED** tree; rose; butter; milk; bottle

People hear with the — eyes; ears; nose; mouth; hands

**SHY** is the same as **BOLD; COY; FRIGHTENED; TIMID; SHINY**

Florida in cotton button grows

1 2 4 16 64 256

eat: bread:: drink: WATER; IRON; LEAD; STONE

If I had 50 per cent more money than I now have I would then have

\$84. How many dollars have I now? .....

ABIDING fortitude; faith; hatred; anxiety; attitude

Chard is a — fish; lizard; vegetable; snake; fruit

etc.

Interpreting intelligence scores. Some intelligence tests, such as the army test, yield a score which consists of a certain number of points. This can be then standardized in various ways just as in the case with special ability tests. The test may be given to a large number of subjects and the average score computed so that any individual's score can be evaluated by comparison with the average. For finer standards the percentile method is often used. The individual scores are arranged in order from best to worst. The best one of all is called the 100 percentile indicating that the subject equals or exceeds in proficiency 100 per cent of the group. Then a slightly lower number of points is computed, such that those attaining that number of points equal or exceed 99 per cent of the group. This score is called the 99 percentile. Similarly, a 50 percentile individual equals or exceeds half the group. The matter may be made clearer by a brief example. (See Table IX.) Suppose that one person makes a score of 28 points in an intelligence test, another makes a score of 29, 2 subjects score 30, 3 subjects score 31, etc., up to the best one, who scores 39. In the third column we see that 2 subjects score 29 or less; 4 subjects score 30 or less, 7 subjects score 31 or less, etc. These last-mentioned figures may now be converted into per cents of the total number of the subjects, namely, 50. These per cents appear in the last column and indicate that 2 per cent of the subjects score 28, 4 per cent score 29 or less, 8 per cent score 30 or less, etc. Putting it another way, a subject who scores 29 equals or exceeds 4 per cent of the group in intelligence; a subject who scores 30 equals or exceeds 8 per cent in intelligence; a subject who scores 31 equals or exceeds 14 per cent in intelligence, etc. So, instead of saying that a subject scores 31

points, we may say that his score is the 14 percentile, meaning thereby that he equals or exceeds 14 per cent of the group in intelligence. This percentile procedure for conversion of test scores is widely used. In many instances we are interested in basing standards on a particular group of individuals, such as freshmen in college or office workers or unskilled laborers. The percentile method is a useful way of expressing the standing of an individual relative to the standard group. Furthermore it makes it possible to compare an individual's standing in one test with his standing in another test. If he is a 75 percentile in one test and a 50 percentile in another, he is obviously superior in the first, although his raw scores (because of the number of test items involved) may not indicate this difference at all.

TABLE IX. ILLUSTRATING THE PERCENTILE METHOD OF INTERPRETING TEST SCORES

RAW SCORE	NUMBER OF SUBJECTS MAKING TEST SCORE	CUMULATIVE NUMBER	PERCENTILE SCORE
28	1	1	2
29	1	2	4
30	2	4	8
31	3	7	14
32	3	10	20
33	6	16	32
34	9	25	50
35	8	33	66
36	7	40	80
37	6	46	92
38	3	49	98
39	1	50	100

Other kinds of intelligence tests yield not a score in points, but a mental age. This is particularly characteristic of the Binet test above mentioned. Certain questions are given for the three-year level, the four-year level, and so on, and by noting what questions a subject answers he is assigned a particular mental age. The usual procedure is then to compute his Intelligence Quotient (I.Q.). This is his mental age divided by his chronological age. If, for instance, his mental age is 12, and his chronological age is 10, his

I.Q. is 120;<sup>1</sup> i.e., he is 20 per cent above the average mentally for persons of his chronological age. If his mental age is 11 years and 3 months and his chronological age 12 years and 9 months, his I.Q. is about 81 (124 months divided by 153 months); i.e., his intelligence is only 81 per cent of what it should be. When dealing with adults this procedure obviously cannot be carried through, for as the same individual grew older, even if he always answered the same questions, his I.Q. would decrease. The standard procedure here is to consider the chronological age 16 for any one who is older than that. A person 50 years old with a mental age of 12 is given an I.Q. of 75; i.e., 12 divided by 16. The assumption is that a mental age of 16 is typical of the average adult. It may be that intelligence stops increasing with age at about the sixteenth year or it may be that it reaches that limit sometime during the teens. The figure 16 has sometimes been called in question, but the tendency is to lower rather than to raise the age at which intelligence stops increasing. Various ages from 13 to 16 have been urged as a basis for the computation of the I.Q. of adults. The most common practice at present, however, seems to be to use 16. The I.Q. then gives an indication of the extent to which the individual's intelligence exceeds or falls short of the average intelligence of persons of his same chronological age or (if he is over 16) of other adults.

#### SUMMARY

Mental tests may be classified according to whether they measure capacity or proficiency. The former deal with essentially innate factors and the latter with acquisitions. The present chapter is concerned with illustrating tests of capacity of the sort that constitute the employment psychologist's stock in trade. They may be further subdivided into tests of special capacity such as attention or memory and general capacity or intelligence. The conventional terminology used in dealing with such tests is justified on the basis of practical convenience. The main consideration is the extent to which the test correlates with the occupational ability which it is desired to predict and its name is in the last analysis irrelevant.

<sup>1</sup> The usual practice is to carry the quotient to two decimal places and then omit the decimal point.



Brief examples are given of tests for motor control, sensory capacity, attention, learning, association, reaction time, space perception, memory, reasoning, decision, ingenuity, and ability to follow directions. There are other less tangible factors, such as emotion, or temperament, for which measures are needed in the practical situation. It is not always a question of what the individual *can* do, but of what he *will* do. The efforts at developing such tests are still in the early stages, but a few illustrations are given.

Notions as to the nature of intelligence vary, but there is apparently some capacity measured by our so-called intelligence tests that gives a person a poorer or better chance for survival in the economic struggle and that makes it possible in certain situations to predict occupational efficiency. This general capacity may be of the abstract type that is ordinarily measured in most tests, it may be of the mechanical type or even of the social type. Illustrations are given of individual and group tests of the abstract and mechanical sort. The scores attained in intelligence tests are usually handled by converting them into percentile scores for the group under investigation or into terms of intelligence quotients.

## CHAPTER V

### MENTAL TEST TECHNIQUE

THE preceding chapter has given a notion of the types of mental tests that are available for a psychologist who is undertaking employment research. As previously mentioned, he needs to know the tools that are available and the proper ones to use on various occasions. But he requires in addition a skill in using the tools and a knowledge of many technical points that must be observed in test administration. A perfectly good plane in the hands of a novice will not produce a smooth plank and a reliable and well-standardized mental test may yield worthless results if not properly administered. The present chapter will be devoted to test technique, with special emphasis on the methods of administration, the devising of test material, and the scoring of results. Most of the principles brought out will be applicable to tests in general, but where this is not the case they will be discussed from the point of view of employment psychology.

#### METHOD OF ADMINISTRATION: INDIVIDUAL VS. GROUP TESTS

There are two methods of giving tests — the individual method and the group method. As their names imply, in the former one person at a time is tested, while in the latter a number of people take the test simultaneously. The individual method involves one examiner for each subject being tested at a given time. In the group method the number of individuals tested by one examiner is limited only by the number of seats and the acoustics of the place in which the tests are conducted. The testing of five hundred persons simultaneously is a common occurrence.

Comparative advantages. Each of these methods has its advantages and disadvantages. In the individual test the examiner is in a position to observe everything the subject does and if anything goes wrong he is able immediately to make the proper adjustment. In a group of people being tested there are some who, in

spite of all precautions taken to make the directions fool-proof and to administer the tests without a hitch, get a bad start or do what they are not supposed to do. Such a simple thing as turning to page 6 after vigorous exhortation by the examiner to "turn to page 4" will frequently occur in a group. Or some subjects will work at such a high pitch of attention that they will fail to see the word "stop" printed in bold-face type. Or if the examiner queries, "does every one understand what he is supposed to do?" some members of the group who do not understand will maintain respectful silence. If the examination, however, is given individually, the examiner will notice if the subject turns to page 6 instead of 4 and will correct the mistake instantly; or if he runs by the word "stop" will immediately call his attention to the fact; and if the subject does not understand the directions he will be more inclined to admit it when not in the presence of other subjects, and, at any rate, in his initial attack upon the test will manifest his lack of understanding. The individual test, then, has a greater certainty that the subject will do what he is told, that he will get a proper start, and hence that the results will be typical of his ability under the prescribed conditions.

A second advantage of the individual test is that it provides more of a "clinical picture" of the subject. In a group test the examiner obtains no data except from scoring the test blank. There are occasions, however, when it is important to observe how the person goes at the test. If he attacks it with zest and apparent effort, his results are perhaps of some value, while if he goes at it listlessly and with apparent lack of interest, this attitude doubtless vitiates the test score, but may be symptomatic of other things with which the examiner is concerned. A psychopathic subject under the pressure of the test situation may manifest emotional disturbances which he would not show under ordinary circumstances. If a certain portion of the test is not marked at all, it is impossible to tell, in the group method, whether the subject overlooked it, misunderstood, was unable to do it, lost interest, became frightened or angry, or had his attention distracted by something else. While this "clinical picture" is usually more important in examinations given to cases of suspected mental disease or mental defect, it is sometimes impor-

tant in the employment problem. The writer was examining a man who had supposedly recovered from shell shock, with reference to employment on a fatiguing job requiring considerable patience and involving rather complex machinery. The man reacted normally at the outset, but in the course of the first test "blew up," cursed the tests lustily, and manifested other psychopathic symptoms. Obviously, it would have been dangerous for him to undertake the work in question and he was given an unskilled job with simple implements outdoors. In a group test it is doubtful just what he would have done and it is certain that the results on his test blank would not have been as illuminating as his remarks. The extraneous reactions of the person during the test are then in some instances of interest and of practical importance.

A third advantage of the individual method is that it permits greater flexibility in the selection of the tests. There are some tests that necessitate material equipment ranging from a picture puzzle up to an electrical device worth hundreds of dollars. In a group test every person must have the same kind of blank or apparatus, and if the latter is expensive it is often unwise to provide a lot of duplicates, especially in the early experimental stages of the project. The natural result is a limitation in the tests that are to be tried out if the group method is used. In some problems, such as selecting clerical workers, this does not seem to be a very serious drawback, but in analyzing some types of vocational ability, such as flying an airplane, it is highly desirable to evaluate rather complicated mechanical techniques. In general the more tests tried the better final selection of tests for an occupation it is possible to make. The individual method affords this greater flexibility in selection.

Over against this array of advantages of the individual test there is only one outstanding advantage of the group test — its economy of time. This is a tremendous advantage, however, in the practical situation. For instance, at Ohio State University in 1919 the army test was given to 6000 students in one day by six examiners aided by a number of assistants. If the test had been administered individually, an examiner working on a reasonable schedule could have finished the job in about a year. In the army something like 100 examiners tested some 1,726,000 recruits within about a year.

It would have taken one man between 600 and 700 years to do this job individually. In the practical situation it is necessary to set the aforementioned advantages of the individual test over against the saving of time and expense of the group test.

There is a scheme that is often used, however, to maintain some of the time-saving of the group test without sacrificing appreciably the advantages of the individual test. This scheme involves the use of a small group — perhaps ten or a dozen. A group of this size may be seated at tables with space between them or in some other fashion so that the examiner by walking around the room can look over every one's shoulder. He may then give almost as much supervision and make almost as careful individual observations as he would in the individual test. After he gives the signal to begin work, he can walk around rapidly and a glance at each paper will tell him whether every one has started correctly and has apparently understood the directions. He can also note whether the subjects turn to the right page or stop at the proper place, and observe numerous other things just as he would in the individual procedure. In all such cases it is possible almost immediately to make the proper adjustments, such as assisting in the finding of the place or giving supplementary explanation where warranted and if necessary allowing extra time to compensate. The examiner can notice, moreover, many individual aberrations in attitude, because with the small group he can give a certain degree of attention to all of the subjects. He will doubtless "spot" any one who is reacting in unusual fashion and observe him more closely. In short, the first two advantages of the individual test may be obtained in the group test provided the group is small.<sup>1</sup>

The other advantage of the individual test above mentioned, namely, the possibility of using more equipment, cannot be obtained in the group without considerable outlay. However, a combination of the two methods is sometimes possible. Suppose that the entire program for each individual involves ten tests that employ printed blanks and two that require technical equipment. It can sometimes be arranged to give the tests involving blanks to the persons simultaneously and then have those same persons return individually for the two tests necessitating apparatus. In testing

applicants for a job, it is often possible to give them a portion of the test in a group and then let them wait while each is given his individual tests. In examining employees, where it causes too much confusion to have each person leave his work twice to be tested, a certain amount of time can be saved by scheduling appointments so that there will always be two persons taking the group part of the test simultaneously. For instance, the first man comes and takes his individual test. Just as he finishes, according to a careful schedule, the second man enters and they take the group part of the examination together. The second man stays on for his individual test and then a similar procedure is repeated with the third and fourth men. Thus it is possible to save time and expense by dividing the tests into those which require apparatus that cannot be duplicated and hence demand individual administration and those in which group administration is feasible.

**Comparative difficulty of technique.** Mention should be made of a further difference in the methods from the standpoint of technique. The individual method usually necessitates a somewhat more skilled or better-trained examiner. The group test is usually somewhat more fool-proof and somewhat safer in the hands of the untrained. This difference is not theoretically intrinsic to the methods. But in the tests that have been devised particularly for individual use, such as the Binet test, the examiner has to use considerable tact and judgment in the course of the examination. In giving directions verbally, much depends on his emphasis and he has to guard against helping the subject with the answers more than standard procedure allows. In reading numbers to be memorized, he has to control the time carefully, and there are many ways in which the novice can vitiate the test results. In the ordinary group intelligence scales, however, the test is often almost self-administering and about all the examiner has to do is to operate a stop-watch and say "begin" and "stop" at the proper moment. The directions are all printed on the blank so that the personal equation of the examiner does not enter. The greater necessity of having a skilled examiner does not, of course, apply to tests actually devised for group procedure, but given individually.

In embarking on a testing program, then, the decision as to what

tests to use will depend somewhat on the ultimate organization with respect to the conduct of examinations. If the methods are to be left ultimately in the hands of persons without psychological training (a condition by no means desirable), it is unwise to introduce any individual tests of the sort requiring particular technique on the part of the examiner. In such a case it is better to adopt group tests or at least tests arranged in as fool-proof form as the usual group test.

**Organization for administration of tests.** Group and individual tests require a somewhat different organization for their administration. The former necessitates, of course, a room large enough to seat comfortably as many as are to be tested. It is desirable to have sufficient space between the subjects so that they will not copy from one another's papers or else provide the test in two forms of equal difficulty and distribute alternate forms to the subjects in alternate seats. In testing a large group it is further necessary to have assistants to aid in the prompt distribution and collection of blanks in order to insure that the subjects do not begin work before they are told nor continue working after the signal to stop. The individual test, on the other hand, needs seating facilities for only one subject, but requires space for whatever technical equipment is used. A room for individual testing often resembles a small laboratory. Usually the examiner can handle the individual test alone, although there are instances where an assistant is desirable to take readings on the apparatus or to make notes of the subject's responses.

#### ADMINISTRATION: METHOD OF TEST RESPONSE

**Oral Method.** The subject may be required to make his response by various methods — oral, written, or performance. As their names imply, the subject may speak his answer, may write it on the paper, or may manipulate the test material in some other way. In the earlier types of test the oral method had some distinct advantages over the written, for it was possible to time the response more accurately and abstract from any error due to differences in speed of writing. In giving a free association test, for instance, in which the subject is started with a word and then gives associated words

as rapidly as they come to him, if he is required to write down these words in succession it is probable that they will come to him faster than he can write so that the result will be no real indication of his speed of association, but will be rather a measure of his speed of writing. In the individual test the actual speed with which the person speaks the words can be timed.

**Written method.** The written response is obviously necessary for the group method. The effect of a group of subjects giving simultaneously their individual free associations would be like unto a confusion of tongues. It would be impossible for the examiner to record the responses of each individual. Of late years the technique of written tests has been modified in one important respect to obviate the difficulty mentioned in the preceding paragraph. Most group tests now minimize the writing of actual words by the subject. He merely has to cross out or underline or use some easily written symbol. For instance, the original "opposites test" consisted of a list of words like:

GOOD.....  
LITTLE.....

and the subject wrote on the dotted line the opposite of the word at the left. This test has been more recently put in the following form:

GOOD is the opposite of: NICE; FINE; BAD; POOR  
LITTLE is the opposite of: LARGE; SMALL; BIG; SHORT

and the subject underlines whichever of the four words in capitals correctly finishes the sentence. With the former type of test, persons who had equal facility in association might score unequally because it would take one of them longer to write "bad" or "big" than it would take another. In the present type, however, there is very little difference in the time taken by various persons to do the underlining. Consequently the test measures their speed of association rather than their facility in motor performance. Where the nature of the test lends itself to this kind of arrangement, the advantage of the oral over the written method of response disappears. There are also test situations in which, though actual words are written, the speed of writing does not introduce a serious error



because the time spent in writing is slight compared with the time spent in deciding what answer to write. For instance, in a test comprising items of this sort:

A E U E U O A E  
 U A U O  
 E E U U O A A E U

in which the problem is to discover the relation of the letter O to the rest of the line that is the same in all three lines, the time spent in writing down the answer "after the second U" is slight compared with the time taken to discover that relation. In such cases the written form of response is as satisfactory as the oral. Inasmuch as the written method is necessary in group tests and these group tests are desirable because of their time-saving, it is fortunate that these modifications in the technique of written responses have taken place.

**Performance method.** In certain kinds of tests it is impossible to use either the written or oral type of response. For instance, in assembling a picture puzzle the subject cannot tell verbally how to do it nor can he write out the method in detail. It is necessary for him to *do* it. Similarly in assembling simple mechanical contrivances in tests for ingenuity, or in performing a series of complex motions in imitation of the examiner, it is necessary to have the subject actually make the motions. In measuring one's reaction time he must press or release a telegraph key when he sees a signal. There have been recent efforts to adapt some tests of this sort to the written form so that they can be given by the group method. Tests of the puzzle type sometimes have pictures of the loose parts numbered and the subject puts the numbers in the proper place on the blank to show where the parts belong. If the examiner touches a series of four points repeatedly in a complex order, the subject, instead of imitating him directly, may write the numbers of the points in the order in which they were touched. However, there will probably always be some kinds of tests which it will be impossible to adapt to written form and there the performance type of response will have to be maintained.

In some instances it has been possible to use the performance method and still administer the test to a group. If the equipment is not too elaborate and is of such a character that an adequate score can be obtained from the finished product, it is possible to provide duplicate equipment and test a group simultaneously. For instance, in a test involving the assembling of small mechanical contrivances — spring clothespin, paper clip, mouse trap, etc. — duplicate sets are provided and the subjects work half an hour. The partially assembled sets are then collected and can be scored at leisure according to a carefully worked-out system. They must be scored, however, and the materials taken apart before another group can be tested.

ADMINISTRATION: TYPES OF TEST RESPONSE

**Free response.** The subject's response to the oral or written test may be either free or constrained. In the former no restriction is placed on the response — the subject writes a "tactful letter" or connects the dots on the page with lines in any way he pleases or gives any words that are suggested by the stimulus words. This type of response is generally difficult to score unless one is merely interested in the time taken to write the letter, to make the marks, or to speak the word. Consequently it is seldom used.

**Constrained response.** The response is most frequently constrained and this may be done in several ways. First of all, it may be constrained by the wording of the question or item. For instance, the subject is given a list of words and required to give the opposite of each rather than merely to give any word suggested by the stimulus word. Or he answers questions of the sort: "arm is to elbow as leg is to what?"

In the second place, the response may be constrained by the location of the answer. This is typified by the "completion" test in which words are omitted from a text and the subject supplies the missing words, as in the following: "In winter the \*\*\*\* is on the ground and the \*\*\*\* blows it into big \*\*\*\*\*." Or the words may be given with certain letters missing like: "c\*w," "ho\*\*e," "d\*ck," and the subject supplies the missing letters. The subject may or may not be informed regarding the number of letters omitted, but

the essential point is that his response is determined by the context and by the location of the answer.

The third way of constraining the response is by having the subject select his answer. He is provided with alternative answers from which he chooses the correct one. The number of alternatives may obviously be 2, 3, 4, or more. The following illustrations are typical:

Good — Bad . . . . . same — opposite  
 A Zulu has TWO; FOUR; SIX; legs  
 Oyster: shell:: banana: TREE; PEEL; SIDEWALK; FRUIT

In the first instance the subject has to indicate by underlining whether the words are same or opposite; i.e., has to choose between the *two* possibilities. In the second case he has to decide between the *three* possible numbers of legs for a Zulu. In the last example he has to choose one of the *four* words that bears the same relation to "banana" that "shell" bears to "oyster."

The most important consideration with reference to the number of alternatives is the possibility of getting the correct answers by guessing. With the two alternatives a person who knows absolutely nothing about the items involved and merely guesses at each will get approximately half of the items correct, just as in throwing a coin a large number of times approximately half of the throws will be heads. Hence, unless some allowance is made, a person may attain a respectable score in such a test and apparently possess ability of the sort involved when this is not the case at all. With the three alternatives the chance of guessing the correct one is somewhat smaller — approximately one chance in three. In such an instance the score attained is more apt to represent the subject's actual capacity, although even here there is some possibility that accident will play into his hands. With four alternatives the probability of making a high score in the test by accident is rather small and with five or six alternatives it is so remote that it is usually disregarded altogether. Tests with four or five alternative answers for each item are probably the most widely used to-day.

## ADMINISTRATION: LIMITATION UPON TEST RESPONSE

**Time limit vs. work limit.** Some limitation must obviously be placed upon the subject's responses in taking a test. He cannot work for an indefinite length of time, nor can he have unlimited material with which to work. Consequently, it is necessary to set either a time limit or a work limit. In the former the subjects all work for a constant length of time — e.g., four minutes — and are graded in accordance with the amount that they accomplish in that four minutes. In the work limit they all finish a constant amount of test material — e.g., selecting 40 opposites — and the number of minutes and seconds required to finish the task constitutes the differential score.

**Time limit preferable in a group test.** The time limit and work limit are equally adaptable to statistical treatment of the results. The time limit, however, is generally to be preferred for a group test. It is possible to have a number of subjects work simultaneously for the same length of time and then subsequently score their individual accomplishment. If the members of a group are required to complete the same amount of test material, it is difficult to obtain a record of the time required by each individual to finish the test. This is sometimes attempted by placing a fast clock where it is visible to all of the subjects and starting them together and then having each as soon as he finishes look at the clock and note on his blank the exact time. In lieu of a clock the examiner may have a series of large cards with the numbers 5, 10, 15, etc., up to 60 for the seconds and 1, 2, 3, etc., for the minutes and display these on a rack. If he carefully follows a watch and changes cards every five seconds, the subjects as they finish can note the time that is displayed. This work limit procedure, however, implies honesty on the part of the subjects. If one wishes to appear to have a better score than he really merits, it is merely necessary to write on his blank a time earlier than the actual one at which he stops. In testing subjects who are willing to cooperate and for whom nothing is at stake, it is perhaps safe to let them record their own times. But in the usual employment situation where a job may be at stake, it is dangerous to trust a person in this way. Unless the test is given individually

so that the examiner himself can measure the time consumed, the time limit is to be preferred to the work limit.

**Work limit feasible with a long test.** It is not usually feasible to have the subjects bring their papers to the examiner as they finish and let him record the time. Most projects involve a group of tests each of which requires only a very few minutes. These may be given in succession, but the time for each must be recorded separately. If a test requires only one or two minutes for its completion, it is obvious that the time taken in bringing the papers to the desk will make a very appreciable increment. Suppose two persons finish simultaneously, but one is in the front seat and the other in the back of the room, the former may get a score of one minute and the latter of one minute and fifteen seconds. This difference of twenty-five per cent will be entirely misleading. This sort of procedure is justified only when the time taken in actually completing the test is very large relative to the time taken in bringing the blank forward and having it recorded so that the latter is negligible. If the test itself takes half an hour the fraction of a minute involved in getting the time record will be insignificant. There is one type of test designed specifically in the light of the foregoing facts — the “omnibus test.” In this type the different kinds of test items alternate throughout rather than appear in separate groups and the only score desired is the total time for all the items. With this sort of test the above procedure is justifiable and it is possible to give the test to persons who drop in at irregular intervals by merely marking on the blank the time they begin and the time they return the paper. In this way it is unnecessary to wait for a quorum before beginning to administer the test.

**Determination of proper limits.** The amount of material selected for a work limit test depends on two things. On the one hand, enough material must be used to give a fair sample of the ability in question. Half a dozen items may not be typical, while 100 may be little better than 75. This all depends on the type of test. On the other hand, the amount of material is somewhat determined by the approximate length of time that can be devoted to the test. It is usually undesirable to include so many items that subjects will require three hours to finish that particular group of items. By

giving a preliminary test to a few individuals, it can be ascertained how many items can be done per minute by the slowest workers and then a proper number selected so that none will need more than the available time.

With the time limit method it is important to determine in advance exactly what limit will be most satisfactory for the material that is provided. The general principle is that the time limit shall be such that the best individual will very nearly but not quite finish the entire test. If many of the subjects finish the test, it is impossible to differentiate between their ability, for one may have barely finished, while another may have had a minute to spare and could have done a considerable number of additional items had they been available. On the other hand, if the best person finishes only half the items there is no need for having the other items on the blank at all. The usual practice is to give the test in a preliminary way to a number of individuals, note the time taken to complete it by the most rapid one, and then select the final time limit slightly less than this figure.

**ADMINISTRATION: GENERAL PRECAUTIONS**

**Standard conditions.** There are a few precautions of a general nature to be observed in giving tests. The examiner has to adapt himself to the conditions available with reference to many details such as the arrangement of materials and equipment, reception of persons to be tested, etc. One fundamental point, however, must be observed. All the subjects must take the tests under standard conditions. A chemical reaction does not depend appreciably on ventilation, room temperature, time of day, external noises, or nervousness of the elements involved. In a psychological laboratory or test room it is altogether different. If some subjects take tests when surroundings are quiet and others take the same tests when a freight train is being made up outside the window, the latter are at a disadvantage and the results are not comparable. The same is true if one group takes the tests at the end of the day when fatigued and another group takes them in the morning when fresh. Likewise if one test room is well lighted and another has illumination of insufficient intensity or has distracting glare, results under the two

conditions cannot be compared. If some subjects use pencils that are too hard and sharp and stick into the paper causing delay, there is a source of error introduced. Psychological experiments reveal the extent to which rather slight changes in environmental conditions influence mental efficiency. Some persons may be able to abstract from or ignore such things, but the natural tendency is to be affected by them. Inasmuch as in a mental test we are attempting to measure one thing at a time, it is desirable to exclude these variables that may influence the results. Consequently, it is of importance to keep the test conditions standard and constant as far as possible.

**Proper attitude.** Another general precaution that it is well to observe deals with the attitude of the examiner and the subjects. It is quite possible for the former to inspire an antagonistic or an alarmed attitude on the part of the latter. A subject who is resentful will probably not do his best and one who is frightened is liable to be somewhat distracted by the emotion. Consequently the examiner should at the outset establish "rapport." This term was used originally in hypnotic technique, but has been aptly applied to mental test procedure. If A is hypnotized by B, he will accept suggestions from B and carry them out, whereas if C tells him to do something the suggestion will not be effective. This is explained by the fact that A and B are "en rapport" and A is more inclined to cooperate with B than with C. Similarly in giving mental tests the examiner should get the subject into this attitude of cooperation or, in everyday parlance, get the subject "with him." Under these conditions the subject will do what he is told, will do his best, and will try to conform to the wishes of the examiner. The establishment of "rapport" calls for tact on the part of the examiner, sometimes an explanation of the purpose of the test project (depending on the intelligence of the subjects) and a general atmosphere of cordiality. It is often well to precede the tests with a few moments of general conversation or with remarks leading up to the matter in hand, gaining the confidence and good will of the subjects and allaying suspicions or fears. Often a "shock absorber" is used for the last of these contingencies. This is a brief test which precedes the others and is not scored at all, but merely serves to get the subjects

accustomed to the test situation. The examiner must adapt himself to circumstances, but whatever they may be should strive to get "rapport" with the subjects and have the whole atmosphere of the examination one of willing coöperation.

There is another point to be observed particularly by the inexperienced examiner. He should himself be thoroughly familiar with the test procedure before administering the tests. If he makes mistakes or has to change his directions after giving them, it is embarrassing, the subjects lose confidence and it is liable actually to vitiate the results. He should rehearse his part if necessary so that there will be no "hitch" in the proceedings.

#### TEST MATERIAL

**Difficulty of material.** In devising material to be used in a particular mental test, one thing that must be considered is the difficulty of the test items. Most tests comprise a considerable number of separate items of the same general sort; e.g., 30 examples of opposites. These should not be made up and used at random, but rather the difficulty of each separate item should be determined. This is usually done by experimenting individually with a number of subjects and measuring the time taken to do each single item. If the results for the various subjects show fair agreement with one another, the average time for an item may be taken as an index of the difficulty of that item.

Assuming that the difficulty of the various items is known, there are two different trends in test construction — to arrange the test so that all of the items will be of approximately equal difficulty (speed test) or to have them increasing rather uniformly in difficulty (power test). In the first of these the interest is in the amount of performance per unit time, while in the second it is in the ultimate difficulty of performance that can be attained. The speed test may be typified by a page of random numbers in which all pairs of adjacent numbers whose sum is 10 are to be cancelled. All such pairs will be of approximately equal difficulty and the number cancelled per minute or per some other constant time interval will be the individual score. Consequently, if one person scores 100 and another 125, it may be stated that the latter is 25 per cent superior



in this sort of performance. Almost any sort of test may be given in this speed form, provided enough items of equal difficulty can be devised. It is most frequently used in situations where one is interested in the subject's alertness or ability to think or act quickly. It intentionally and avowedly puts a premium on speed.

The power test may be typified by a number completion test in which a series of numbers are given and the subject is required to complete the series. The items may start with relatively easy ones like:

1 2 3 4 5 6 7 .... ....

and lead up through gradually increasing degrees of difficulty to items such as:

2 4 8 3 9 27 4 .... ....

Power tests are usually given with a time or work limit, but the temporal aspect is not regarded with as much concern as in the speed test. If a time limit is set it is usually such that the subject will get about as far along in the items of increasing difficulty as he would if he had unlimited time. While he might do a few more items if he had opportunity to take the blank home overnight (subjects occasionally make this request), he would not do very many more and the number of items he passes under the test conditions is a pretty fair indication of his proficiency in this particular sort of task. The power test is most often used in situations in which interest is not in a person's intellectual alacrity, but rather in his ultimate possibilities of intellectual attainment.

There is a popular misconception that should be cleared up in this connection, namely, that the speed test is not a "fair test." The subject states that if more time had been allowed he would have been able to do better, and that he has known persons who would be very slow in thinking out items of this sort, but who were nevertheless economically and socially successful. Of course the subject might do more with unlimited time — and so would his competitors. But the purpose of the speed test is not to find out how much he can do at leisure, but how much he can do per unit time. As mentioned previously the tests are so constructed that very few persons will

finish, in order that the scores may scatter over a considerable range. To be sure, our best work in daily life is not done to the time of a stop-watch, but it is true in general that the brighter minds work not only better but more rapidly. After all, the "fairness" of a test depends on whether it may validly be used in predicting some correlated capacity. If scores in power tests are more closely correlated with proficiency in clerical work than are scores in speed tests, the former will be fairer to use in selecting clerical workers and *vice versa*. As a matter of fact, statistics show that the abolition of time limits would be in many cases disastrous, for there are definite tendencies for those who are proficient in tests which emphasize speed to make better messenger boys, clerical workers, salesmen, engineers, and to rise in general to occupations on the business or professional level rather than on the level of unskilled or semi-skilled labor. Where tests devised for a practical purpose, such as predicting engineering aptitude, were given with and without time limits, their diagnostic value was greater in the former case. (606, 275.)

**Arrangement of test material.** The usual procedure in assembling tests is, as implied above, to group together items of a given sort. A project such as an intelligence test may comprise several different kinds of items which it is desirable, however, to evaluate separately. This is obviously facilitated by grouping them so that, for instance, the "attention test" and the "memory test" are entirely separate. Each group of items is generally preceded by the directions or instructions for dealing with those items. Usually the blank is so arranged that one set of items occupies a page and with its directions forms more or less of a unit. In the Army Alpha intelligence test, for instance, the first page comprises 12 items for which verbal directions are given and is labeled test 1; on the next page is test 2 comprising 20 simple arithmetical problems with directions printed at the top of the page; the next page constitutes test 3 on practical judgment comprising 16 questions with three alternative answers to each — the subject to check the best answer as explained in directions at the top; in test 4 are 40 pairs of words which the subject checks to indicate whether they are synonyms or opposites; and the remaining 4 tests are presented in similar fashion.

Each page lists test items of a separate kind with directions at the top of the page. This is typical of most scales or groups of tests used in experimental work — separate administration of different kinds of test items.

Omnibus tests, however, depart from the foregoing arrangement — and are designed to facilitate test administration. Instead of arranging all of the items of a given sort so that they occur together in a single test to be followed by all of the items of another kind grouped together, each test with its time limit of a few minutes, the items of the different types are intermixed in one way or another with a single time limit or even a work limit for the whole. For instance the army test (just mentioned) or parts of it have been recast into various omnibus tests. A typical one starts with three arithmetic problems followed by three practical judgment items followed by three disarranged sentences and so on, returning then to three more arithmetic items, three more practical judgment items, etc. Whereas in the original army form the subjects were allowed 5 minutes for the arithmetic test,  $1\frac{1}{2}$  minutes for the practical judgment test,  $1\frac{1}{2}$  minutes for the opposites test, etc., in the omnibus form above mentioned they are allowed 30 minutes for the entire scale. The choice of three successive items of a given sort is arbitrary. It might have been only one or it might have been 10. In some instances the experimenter is interested in providing very quick shifts of attention from one sort of thing to another in order incidentally to measure this factor as well as the general ability manifested in the test itself. Sometimes he is more concerned with getting the subject well started on one sort of item before the shift occurs in order to note his ability to change his "set" after it is well established. The items may even be given in a random order. In the omnibus test the directions or explanation for all the kinds of items involved must necessarily precede the test proper.

The omnibus test, like that with the items grouped, may use items either of equal difficulty or of increasing difficulty. When all the items of a given sort are approximately equal in difficulty, the test is called a "cycle omnibus" test. When the items of each sort increase in difficulty throughout the test — i.e., each item is more

difficult than the preceding item of that sort — the test is called a "spiral omnibus" test.

**Alternative material.** In making up test material it is well to devise additional items at the outset in order to provide alternative test blanks. If the original blank is in use for a considerable time, it is quite possible that a copy will get outside so that some persons will have access to it before taking the tests. Moreover, those who are examined will remember some of the items and discuss them with friends. Every one engaged in a test project of any magnitude feels occasionally that some of the persons who come in to be tested are not as naïf as should be expected. A subject not infrequently beams with pleasure when recognizing items that are familiar and on which he has been "primed." When this situation arises, it is desirable to have another test blank involving different items, but of the same difficulty as the first. Other persons can then take the tests without profiting by any previous information they may have received, yet their results will be directly comparable with those of the persons who have been tested previously. Perhaps the best way to arrange these two blanks or "forms" is to provide about twice as many items as are necessary for one form, when devising and determining the difficulty of the original items. Then if the figures representing difficulty are available for all items, it is comparatively simple to select two groups of items with the same total difficulty. These two forms of the test are likewise of value in cases where it is necessary to test large groups of individuals crowded together so that there is danger of their copying one another's papers. The blanks may be distributed in such a way that subjects in adjacent seats have different forms. In fact most of the larger test projects issue a given test or scale in two or more alternative forms to provide for the contingencies above indicated.

**Sensitivity.** The test material should be selected with a view to sensitivity. A sensitive test is one that gives a considerable range of test scores with the group studied or that reveals marked individual differences in performance. If every one taking the test scores either 29 or 30 points, it is not considered a sensitive test, whereas if some individuals score as low as 10 points and some as high as 80, the test differentiates clearly between the various

subjects. In securing this sensitivity there are two things to consider. In the first place, the test should have a lot of items or increments. Suppose a test involves only 3 items, it will be possible for the subjects to score 0, 1, 2, or 3 points. The best that can be done is to divide the subjects into four degrees of ability. In the second place, the items should be selected so as to be differential with the group studied. It is possible to have the items all so easy that every one can do them as rapidly as he can write or make the appropriate marks. If, for instance, a group of college students are given a test of the order of  $2 \times 3$  or  $3 + 5$ , they can do the problems as rapidly as they can write the answers, and it is probable that they will all make approximately the same score and hence the test will not be sensitive. On the other hand, a group of persons of low intelligence may be given questions that are so difficult that none of them will be able to do more than one or two. If, however, the difficulty of the questions is neither too little nor too great for the individuals being examined and if there is a sufficient number of questions, the test will be sensitive and reveal the desired individual differences.

#### TEST INSTRUCTIONS

**Standard instructions.** The instructions given the subjects are almost as important as the test material because they must insure that the subject will do what the examiner actually wants him to do. Perhaps the most important point about instructions is that they must be kept standard or constant. If one person is told to do one thing and another person told to do something else, obviously their test results are not comparable. If one blank says, "Work as fast as you can," and another says, "Make no mistakes," quite different attitudes will be evoked and an altogether different emphasis on speed or accuracy given to the subjects. The second blank may show greater accuracy than the first, not because the individual using it is naturally more accurate, but because he is told to be more accurate. If one subject is instructed to complete every item before passing to the next and if another subject is told to skip any items which he cannot solve in a few seconds, the first may spend half the test period on some single item which he finds difficult,

while the other may make a much higher score simply because he selects the items which he can solve easily. The emphasis on factors like these must be determined by considering whether the examiner is more concerned with speed or with accuracy in the particular problem for which the test is to be used. But the essential point is that once the instructions have been determined upon, they must remain constant for every one who takes the tests. Sometimes, of course, supplementary explanation is given if the subject does not understand the original instructions. It is well to have this standard likewise so that no subject will be given any unfair advantage due to some implication in the wording. As a matter of fact, ideal instructions will need no supplementing, at least with adults of normal intelligence.

**Clarity of instructions.** Another requisite of the instructions is their clarity. If they are ambiguous or incomplete so that the subject does not understand exactly what is wanted, they fail of their purpose. It is not safe for the examiner to compose the instructions and employ them at once. It is highly desirable to try them out on a few persons, preferably of the type with whom the tests are to be used. Instructions that seem absolutely fool-proof to the one who writes them will frequently have some point which can be misinterpreted or some contingency that is not covered. If the subject is told, for instance, to "mark the correct word in each line," he may underline it as the examiner intended or may waste his time making elaborate rectangles about the words. He may work up and down the page, although it was assumed that he would do the obvious thing and work across. If he is told to cancel the vowels he may be in doubt as to whether *w* and *y* are respectable vowels. Any number of minor points of this sort will come out in using test instructions. Hence it is well to give them to a small experimental group and note any questions that are asked and any uncalled-for performance that results. The instructions can then be modified accordingly before putting them into practical use.

In insuring the clarity of the instructions it is necessary to consider the general mental level of the persons who are to read or hear them. The vocabulary for persons of low intellectual status must necessarily be simpler than that for persons higher in the scale.

Statistical studies reveal a greater incidence of polysyllables and long sentences in "high-brow" magazines than in those of the "garden" variety. Persons of lower status are likewise apt to require more detailed explanation. For instance, a group of college students, if given a page of numbers in random order and told to "cross out every pair of adjacent numbers whose sum is 10," will probably be able to do it, while a group of unskilled laborers will become paralyzed or profane. It will be necessary to tell the latter: "Wherever you see two numbers side by side that would give 10 if you added them together, draw a line through those two numbers. Remember that they must add up to 10 and that they must be side by side with no other number between." A safe rule in devising instructions is to consider the lowest person who is apt to take the test and step them down to his level. The others will perhaps be bored, but this will not vitiate their results, and it is better to play safe and to insure that even the poorest one in the group understands what he is to do.

**Form of instructions.** The actual form of the instructions naturally varies with the test involved. However, most instructions embody three parts — explanation, illustration, and practice. Some test material is usually presented to the subject while explanation is made as to what is to be done. Then this material is marked by the examiner by way of illustration or else these or additional examples already marked are presented for study. Finally, further unmarked items are given for practice before beginning the test proper. While the subject may think that he understands the test from looking at the illustrations, he may find it a different matter when he comes to work out practice items himself. If he accomplishes these latter, it is then certain that he understands what he is to do in the actual test. The following excerpts from the directions preceding a group omnibus intelligence test illustrate these three stages of explanation, illustration, and practice.

Inside this booklet you will find a lot of things to do. Samples of the different things to be done are given below, along with a few examples on which you can practice. You will be given plenty of time to study the directions and do the practice examples. These do not count as part of the test but are merely to make sure that you learn to do each kind of problem correctly.

1. GOOD is the opposite of: EXCELLENT; CHEERFUL; BAD; WRONG; TRUE.
2. LITTLE is the same as: SMALL; COARSE; PRODIGIOUS; FEEBLE; IMMENSE.

Underline one of the last five words in each line that makes the best sentence. If more than one answer seems correct underline the one that is the most nearly the same or opposite according to specifications. Mark only one in each line like this:

1. GOOD is the opposite of: EXCELLENT; CHEERFUL; BAD; WRONG; TRUE.
2. LITTLE is the same as: SMALL; COARSE; PRODIGIOUS; FEEBLE; IMMENSE.

Do the following problems for practice:

1. THICK is the opposite of: HEAVY; LARGE; THIN; SMALL; NARROW.
2. SHY is the same as: BOLD; COY; FRIGHTENED; TIMID; SHINY.
3. CARELESS is the opposite of: NEGLIGENT; UNEASY; ANXIOUS; UNCONCERNED; CAREFUL.

. . . . .

1. a eats wood cow grass
2. birds swim feathers have all

The words "a eats wood cow grass" in that order do not make a sentence but they would make a sentence if put in the right order, only there would be one word left over. The sentence would be "a cow eats grass" with the word "wood" left over. The thing to do is to cross out this extra word "wood," like this:

1. a eats ~~wood~~ cow grass

The words "birds swim feathers have all" would make a sentence if put in the right order, "all birds have feathers" with the word "swim" left over. The thing to do is cross out "swim" like this:

2. birds-~~swim~~ feathers have all

Do the following problems for practice:

1. dogs climb meat eat



2. Florida in cotton button grows
3. ocean house in live fish the

.....

- |    |    |    |    |   |   |    |
|----|----|----|----|---|---|----|
| 1. | 2  | 4  | 6  | 7 | 8 | 10 |
| 2. | 32 | 20 | 16 | 8 | 4 | 2  |

Each number is derived in a certain way from the numbers coming before it. Study out what this way is. You will find in each problem one extra number that does not belong there. Cross it out like this:

- |    |    |               |    |              |   |    |
|----|----|---------------|----|--------------|---|----|
| 1. | 2  | 4             | 6  | <del>7</del> | 8 | 10 |
| 2. | 32 | <del>20</del> | 16 | 8            | 4 | 2  |

Do the following problems for practice:

- |    |    |    |    |    |    |     |
|----|----|----|----|----|----|-----|
| 1. | 22 | 24 | 26 | 28 | 29 | 30  |
| 2. | 13 | 12 | 11 | 10 | 9  | 7   |
| 3. | 1  | 2  | 4  | 16 | 64 | 256 |

.....

1. sky: blue:: grass: TABLE; GREEN; WARM; BIG.
2. Locomotive: train:: horse: BICYCLE; HUB; BUGGY; BAGGAGE.

The first word "sky" is related to the second word "blue" in the same way as the third word "grass" is related to one of the words following it. You are to underline the word that is related to the third word as the first two words are related to each other. In this example "sky" is related to "blue" as "grass" is related to "green" because the sky is colored blue and the grass is colored green. Therefore "green" should be underlined like this:

1. sky: blue:: grass: TABLE; GREEN; WARM; BIG.

In the second example, "locomotive" is related to "train" as "horse" is to "buggy," for a locomotive pulls a train and a horse pulls a buggy. Therefore "buggy" should be underlined like this:

2. locomotive: train:: horse: BICYCLE; HUB; BUGGY; BAGGAGE.

Do the following problems for practice:

1. bird: sings:: dog: FIRE; BARKS; SNOW; FLAG.
2. eat: bread:: drink: WATER; IRON; LEAD; STONE.
3. arm: elbow:: leg: FOOT; KNEE; SHOE; KICK.

In some cases, of course, the test is so simple that elaborate instruction is unnecessary — for instance, "Solve the following arithmetical examples," or, "cross out every letter A on the page." Sometimes the practice examples are omitted, but this is unfortunate and particularly so in tests of a motor character. If the subject is actually manipulating apparatus there is bound to be a very appreciable effect of practice in the way he holds his hands or the implements, and it is always wise to let him acquire this initial adaptation to the test conditions before taking his final record. With the more intellectual types of test this may not be quite as serious, but it is safer nevertheless to give initial practice. Theoretically the function should be practiced until further experience makes no improvement and then the record should be taken. This is usually impracticable and the examiner has to set the limit arbitrarily, governing himself by the time available, by the apparent difficulty of the task, and, if feasible, by the rapidity of progress revealed by a few subjects who take the test repeatedly.

**Printed vs. oral instructions.** Instructions may be oral or printed. Where the subjects are working with test blanks, it is current practice to print the directions on the blank. This has the advantage of eliminating the personal equation of the examiner. The subjects examined at various times get exactly the same wording with no difference in the verbal emphasis. With the verbal instructions one examiner may say, "*Work as fast as you can without mistakes,*" while another may say, "*Work as fast as you can without mistakes,*" and thus evoke quite different attitudes on the part of the subjects. In some cases, of course, printed directions are undesirable. Some subjects cannot read, but can take performance tests and must obviously have the instructions given verbally. Sometimes there is a limitation on printing or mimeographing service so that it is necessary to economize by omitting the printed directions. Sometimes the verbal method is used so as to prevent the subject from working ahead in the blank prior to the signal, although this difficulty can usually be avoided by arranging the blank so that when working on one page the adjacent page is upside down. If the oral instructions are to be used, effort should be made to keep them as constant as the written ones. Most examiners have the

instructions written and actually read them from their copy or have them memorized and give them verbatim.

#### INCENTIVE

**Maximum incentive.** Incentive is a factor that must be controlled. This may often be done through the test instructions and hence it is discussed in the present connection. If incentive is not controlled, it simply introduces another unnecessary variable and this is contrary to scientific method. If a chemist is studying the relation between the pressure and the volume of a gas, he does not let the temperature vary at random, but keeps it constant so as to determine the relation between the other two variables. Similarly, a psychologist studying the relation between intelligence and vocational aptitude tries to stick to those two variables and keep other things constant. So if two persons take the same test and one does the best he can and the other does not try, we have immediately introduced another variable. Their scores may be altogether different, although they have, perhaps, the same actual ability. Incentive, therefore, should be a constant rather than a variable and the only practical way of keeping it constant is to keep it at a maximum. Under these latter conditions we can say that one subject makes a certain number of points when he is doing the best he can and that he is superior to another who is likewise doing his utmost.

**Securing incentive.** It is often possible to obtain this incentive by emphasizing in the instructions the importance of doing well. The exact statements used in introductory explanation of the purpose of the tests will vary with the circumstances, but the final statement that "It is important for every one to do his best" is usually quite effective. In testing applicants for a job, incentive, of course, will take care of itself, because naturally they realize that their score may have something to do with their being hired. In testing employees for research purposes the problem is more difficult. It may be that there is a possibility that the tests will be used for promotion or readjustment of some sort and that it is desirable to let the subjects know this fact. Sometimes there may be an appeal to their pride, to the effect that "We are standardizing these

tests and we want to find what people who are already on this job and making good can actually do in the tests." With more intelligent subjects it is sometimes wise to explain the actual research problem and to enlist their coöperation in a scientific experiment. Occasionally competition may be used as a motive, such as the statement that so-and-so "cleaned up on this test, now see what you can do." In a small group, if after one test the subjects compare notes, such as "how far did you get," and this can be permitted without danger that any one will work overtime, it may serve as an additional motive for the following tests. Competition with one's self is also effective at times. If a test comprises several parts, the subject may be urged in the second part to see if he can beat his record in the first. In individual tests favorable comment on results of a test will often motivate the subsequent tests. The particular sort of incentive that will prove most effective will depend on the type of subjects, the test situation, and the nature of the test. The examiner must adapt himself to these and strive for some effective means of keeping incentive at a maximum. Some final statement in the test instructions to the effect that it is important to do one's best is usually desirable.

#### SCORING OF TESTS

**Unequivocal scoring.** In devising tests consideration should be given to the possibility of unequivocal and simple scoring of the results. The first of these is in the interest of reliability and the second in the interest of time-saving. The unequivocal character is necessary in order to insure that when the tests are scored or administered by various individuals comparable results will be obtained. If it is necessary, for instance, to determine whether the answers to certain test items are good, average, or poor, different examiners will doubtless differ in their judgment. The personal element will enter and different persons scoring the same test blank will obtain a different total. If, however, the items have each a single correct answer, all examiners will obtain exactly the same score for a given subject. Hence it is desirable, whenever possible, to have the items of the single-answer type whether this answer is given verbally, written on the blank, or selected from a list of alter-

native answers. Situations may arise, however, in which the answer must be somewhat equivocal. Even here, it is possible to devise a scheme for making greater uniformity in the results obtained by different examiners or scorers. Suppose it is desired to give an opposites test providing the subjects with a list of words and letting them write after each one its opposite. The subjects may make a considerable variety of responses to a given word, and it is then necessary to determine which responses deserve credit. This is sometimes done by standardizing the items with a large number of subjects and considering the most frequent responses to a given item as correct. In an actual instance, as the opposite to the word "happy" 29 subjects gave "unhappy"; 43 gave "sad"; 6 "sorrowful"; 4 "miserable"; 1 "unfortunate"; 1 "disconsolate"; 2 "sorry"; 1 "discontented," etc. On the basis of these results it was decided to give full credit for "unhappy" and for "sad"; to give half credit for "sorrowful" and for "miserable"; and to give no credit for other words in the subsequent use of this item. In similar fashion standards were obtained for all of the other words used in the test. Then it was possible for different persons to use the key-list of full and half credits for each word and thus to get identical results when scoring the same test blank.

There are other test situations in which the qualitative character of the scoring is still more in evidence. If the subject copies a geometrical figure or writes something to exhibit his own handwriting, recourse is had to some form of rating scale. This consists of a series of specimens of the geometrical figure in question or of handwriting, ranging from very poor quality to very good. These specimens have been standardized and each assigned an appropriate number of points. In grading the test, then, the scorer compares each item in question with the specimens in the scale, determining which of the latter the former most resembles and assigning it the corresponding number of points. With a little practice in the use of such scales fairly reliable results can be obtained. All of the arguments, however, are in favor of the entirely objective and unequivocal type of score wherever the test can be adapted to this form.

**Ease of scoring.** The ease of the scoring from the standpoint of clerical work is especially important in written tests. In the oral

test the examiner usually notes the scores on the items during the progress of the examination, but in tests given to large groups the clerical work of subsequently scoring the blanks mounts up tremendously unless effort is made to simplify the process. One fact that makes for great simplification is that in the printed blank it is possible to have the answers in the same location on each blank. If the answers are to be written, dotted lines or brackets may be provided, thus insuring the location of the answer. If the response consists of crossing out or underlining something, its location is already determined. In the former case, if the answers can be arranged in a column down one margin, a card containing the correct answers likewise in column may be aligned alongside and the two columns easily compared. If the answers are simple symbols such as  $x$  and  $o$ , it is often easier to memorize the sequence. This may be facilitated by arranging the items in such a way that the correct symbols occur in rhythmical sequence, such as "xooxooxoo," so that the subject will not discover the rhythm, but in an order simple enough to enable the scorer to memorize the correct order easily. Even with answers which consist of a list of words or numbers or letters, it is often rather simple to memorize them and this frequently takes place incidentally after the key has for a time been used in correcting the blanks. In case the test response consists of checking words or symbols at particular places on the blank, the correcting can be greatly facilitated by the use of a stencil. A sheet of transparent material such as celluloid is placed over a blank in order to mark on this stencil with india ink the correct places. The stencil can then be aligned over a blank that is being corrected and it is easy to note whether the marks on the blank correspond to those on the stencil.

There are various other minor points which facilitate somewhat the scoring or statistical treatment of tests. It may be desirable to have the lines numbered, provided there is one test item to a line. If there are several to a line, the cumulative total from the beginning of the test to the end of the line in question may be indicated at the end of each line. This will save a few seconds in determining how many items have been attempted after the correct or incorrect ones have been checked. Sometimes it is well to print at the end of

each line the number of items that should have been marked in that line. In case it is undesirable for the subject to know this key, it may be concealed. For instance, if the test consists of numbers or letters that are being cancelled, the last number in the line may actually be the key that tells how many should have been marked or the last letter may represent the correct number in code. It sometimes facilitates matters to have the location of the answers staggered in two columns so that the odd-numbered answers are in one column and the even in another. This is of value where it is desired to total the two separately in order to check the reliability of the test by correlating one half of the test with the other (cf. p. 137). Various other devices in typography or arrangement will often occur to the person devising some specific test.

**Scoring speed and accuracy.** After a test has been corrected one of the most serious problems is to determine what shall constitute the final score. The subject may omit some items; he may get others wrong. The omissions are not usually considered as serious a problem as the errors. Unless specific instructions have been given to omit no item and unless the subject has very patently skipped around and tried to pick the easiest ones, an occasional omission is overlooked and emphasis placed on those actually attempted. In cases where the test consists of finding certain things (as in cancelling A's on a page) the omissions may be counted as errors or else an arbitrary formula devised to weight them. This problem, however, arises in only a limited number of tests, whereas the problem of speed and accuracy is present in a majority of mental tests. There are three ways in which the problem of speed and accuracy score may be handled. In the first place, errors may be neglected and speed alone or number of items correct constitute the sole score. This is reasonably satisfactory when the errors are relatively few. In many kinds of tests the subjects will make comparatively few mistakes, if properly instructed, perhaps not over five per cent. If this is true of all the subjects, it is reasonably safe to neglect the errors.

In the second place, it is sometimes feasible to score only the accuracy or quality of the responses and to neglect the speed. This is to some extent true of the "power" tests (*supra*) in which a

rather liberal time is given for the test or in which every one finishes the test and little account is taken of the time consumed. The answers are scored then entirely on the basis of their quality or accuracy.

In the third place, speed and accuracy may be combined into a single score. This may be done either arbitrarily or statistically. Perhaps the most usual practice is to penalize the subject a certain amount for each error and subtract the total penalty from the total number of correct items. The examiner simply uses his judgment in determining the penalty, deciding whether in the particular situation for which the test is to be used mistakes are very serious, and if they are, making the penalty severe. In such cases it is wise to score a considerable number of blanks, using various degrees of penalty and study the results carefully to see the relative standing of the subjects with the different penalties.

In some types of test the weighting of these factors is more obvious. In the type in which the subject chooses between two alternatives, as has been already suggested, there is an approximately even chance of getting the correct answer by guessing just as in tossing a coin there is an even chance of getting "heads." If there are 100 items and a subject knows absolutely nothing about them but simply marks them at random, he will get approximately 50 correct, while a subject who tries to work them out but goes slowly and painstakingly may not do more than 30, but get these 30 correct. The score of "number correct" will then be entirely misleading, for the first man ought to score zero. This situation is usually met by scoring the number right minus the number wrong. The argument is that the man who guesses on all of the items, and has 50 right and 50 wrong, will score 50 minus 50 or 0, while the man who does 30 correctly and makes no mistakes will receive 30 minus 0 or 30. This seems fair. Or suppose the second man actually knows 30 items, but does not know the other 70 and guesses on them in addition to doing the 30 that he does know. He will then get about 65 correct — the 30 he knows plus 35 or half of those at which he guesses. He will likewise have 35 wrong — half of those at which he guesses. His score will be 65 minus 35 or 30, which is what he deserves for the 30 items he actually knows. Although this method



of scoring the two-alternative test is widely used, it has certain shortcomings. In the statement that if items are marked at random there will be approximately 50 per cent correct, the emphasis is on the "approximately." Many of us have encountered a situation where the pennies within a limited time simply did not match according to theoretical expectation. The laws of probability merely insure that if a large number of people guess at the items the people who get half correct will numerically exceed those who get any other number correct, but this former group is by no means the majority. It is quite similar to tossing coins. Suppose ten coins are tossed a large number of times. Five heads and five tails will be thrown in the long run more often than any other combination, but there is also a possibility of other results. Four heads, six heads, or even two or three heads may occur sometimes, although less frequently than the five heads. Exactly the same thing applies in guessing at test items where there are two alternatives just as there are two sides to a coin. It is possible to compute from the theory of probability what is to be expected in the long run. Suppose that a test contains 10 items (such a brief test would, of course, not be used in the practical situation). It is possible to compute what per cent of the subjects who guess at the items will correctly guess 10, 9, 8, 7 items, etc. These percentages are given in the first part of Table X. For instance, 0.1 per cent of the subjects will in the long run get all the ten items correct, 1 per cent will get 9 of them correct, etc. Similarly, if the test comprises a more reasonable number of items such as 50 (cf. second part of the table), 0.1 per cent of the subjects will get 36 of them correct, 0.2 per cent will get 35 correct, etc. There are still smaller percentages which do not appear in the table for more than 36 or less than 14 items. In both instances it is to be noted that there are more subjects who are apt to get just half of the items correct than there are who are apt to make any other score. However, these subjects by no means constitute the majority — in the first instance they are about 25 per cent and in the second 11 per cent of the group. Obviously, if we score the test according to the number correct, some subjects by mere guessing will get a fairly high score. Some allowance for this must be made. If we make the usual allowance by scoring the number right minus

TABLE X. PROBABLE PERCENTAGES OF INDIVIDUALS GUESSING CORRECTLY VARIOUS NUMBERS OF ITEMS IN TESTS OF TWO-ALTERNATIVE FORM

TEST OF 10 ITEMS			TEST OF 50 ITEMS		
Number of items correct	Number right minus number wrong	Per cent of individuals	Number of items correct	Number right minus number wrong	Per cent of individuals
10	10	0.1	36	22	0.1
9	8	1.0	35	20	0.2
8	6	4.4	34	18	0.4
7	4	11.7	33	16	0.8
6	2	20.5	32	14	1.6
5	0	24.6	31	12	2.7
4	0	20.5	30	10	4.2
3	0	11.7	29	8	6.0
2	0	4.4	28	6	7.9
1	0	1.0	27	4	9.6
0	0	0.1	26	2	10.8
			25	0	11.3
			24	0	10.8
			23	0	9.6
			22	0	7.9
			21	0	6.0
			20	0	4.2
			19	0	2.7
			18	0	1.6
			17	0	0.8
			16	0	0.4
			15	0	0.2
			14	0	0.1

the number wrong as indicated by the second column in each section of the table and call zero all the scores that are by the computation negative, we improve matters considerably. More than half the group thus receive their deserved score of zero, but even then there are some who make rather high scores. In the 10-item test, for instance, 4 per cent of the individuals will make a score of 6, whereas they know nothing about the items; in the 50-item test 4 per cent will score 10 points. This method of scoring still tends to give some individuals a higher score than they deserve.

Probability may likewise work in the other direction and cause some persons to get a lower score than they deserve. It is common practice to make the subject mark all the items of which he is sure

and guess at the rest — otherwise the test might merely reflect his degree of caution. Suppose, however, that two persons know exactly the same number of items and one marks them and stops, while the other marks them and then guesses at the rest. From the sort of thing shown in the preceding table it is evident that there is some chance that the individual who guesses will make more than half of his guesses incorrect, thus reducing his actual score on the ones he knows. An instance is reported of a subject who marked 14 items and said that was his limit. (116.) The examiner noted his answers and there were 13 of them correct, but he told the subject to guess at the rest. The final result was a score of zero because in the end there was a sufficient preponderance of wrong guesses to nullify the 13 correct answers. An experiment was conducted in which it was assumed that various numbers of subjects knew certain groups of items of a 60-item test and coins were tossed for the remaining items to indicate guesses and then the final right minus wrong scores computed. If, for instance, the subject was assumed to know 50 items, 10 coins were tossed for the remaining unknown items and heads counted as a correct guess. If there were 7 heads the subject's final score was 57 minus 3 or 54. These right minus wrong scores were compared with the true scores — i.e., the number which it was assumed the subjects actually knew. It is not worth while to present the results in full, but one or two instances may be cited. It was assumed that 8 subjects actually knew 22 items and guessed at the remaining 32. When their final scores were computed they ranged from 16 to 28. In other words, one subject scored 6 points less than he deserved while another scored 6 points too much. For those who actually knew 24 items, the final scores ranged from 14 to 28 — some receiving a very distinctly different score from that merited. Similar figures might be cited for other numbers of actually known items. The point is that when the subject knows some items and guesses at others, the guessing will not always even up on the unknown items, but may either raise or lower appreciably the score based on the items actually known. It is unwise in such a test to instruct the subjects not to guess, because it is impossible to tell whether all the individuals follow the instructions. The only possibility is to make them guess on all items,

but this has the disadvantages just mentioned. Consequently, when we consider the matter from the standpoint of scoring, the two-alternative type of test has distinct shortcomings. The best that can be done is to score number right minus number wrong, but even this procedure may produce misleading results. Inasmuch as it is usually almost as easy to devise a test with four or five alternatives as to devise one with two alternatives, it is well to dispense with the latter as far as possible.

The best scheme for combining speed and accuracy involves the evaluating of them statistically with reference to the situation in which the test is to be used. Suppose the test is devised for predicting ability in clerical work. As suggested previously, the test must be ultimately evaluated by comparison of test scores with an occupational criterion, namely, actual clerical ability as obtained from production figures or estimates of the office managers. Given this criterion it is possible to correlate with it speed in the test and also accuracy in the test. It can then be determined which is the more closely correlated with clerical ability or which is the more valuable in predicting it. Moreover, it is possible by the technique of partial correlation to determine the best weighting for these two factors. This technique has already been mentioned (p. 55, *supra*) and will be discussed more fully in Chapter IX. Not only are speed and accuracy related in some degree to the criterion, but they are related, perhaps inversely, to each other. It is necessary to determine what the relation of each to the criterion would be if the other were eliminated or kept constant. For instance, if a number of subjects could be obtained who had all exactly the same speed, it could be determined to what extent accuracy correlated with proficiency in clerical work for this limited group; and if another group could be found all with the same accuracy, the correlation of their speed with the criterion could be computed. It is seldom possible to find a group of subjects like this who are constant in either speed or accuracy. It is possible, however, by the mathematical technique above mentioned to obtain the same result from the actually available data. When these partial correlations are found — i.e., the intrinsic relation of speed and of accuracy to the criterion with the other factor constant — it is possible to determine exactly how

much importance or weight should be attached to each. When the speed and accuracy are weighted according to this procedure, the combined scores will correlate more highly with the criterion than if they are weighted in any other fashion. This can be shown theoretically or empirically. In an actual case the correlation coefficient between speed in the test — i.e., number of items completed — and the criterion is .60; the correlation between accuracy in the test — i.e., number of mistakes — and the criterion is -.50; while the correlation between speed and accuracy is -.20. This indicates that those who accomplish the largest number of test items tend to be most effective in the job and *vice versa*, while those who make the fewest mistakes are likewise most effective in the job and those who do the greatest number of items tend to make somewhat fewer mistakes, although this last relation is not very marked. Application of partial correlation technique indicates that the best scoring formula is:

$$\text{Criterion} = \text{Number right} - .76 \times \text{Number wrong.}$$

In other words, if each correct item counts 1 point, we should penalize the subject .76 of a point for each mistake. If the individual blanks are now scored by this formula with speed and accuracy weighted in this fashion, these weighted scores correlate with the criterion to the extent of .71, which is considerably better than the correlation of .60 which was obtained with speed alone. Hence weighting the two variables in this fashion materially improves the prediction of the criterion on the basis of the test scores.

#### RELIABILITY OF TESTS

**Correlating two forms of the test.** The test score is only an approximate measure of the ability in question. It is impossible for an ordinary test to be so devised that it will conform absolutely to all of the principles described in the preceding part of the chapter. Slight differences, in difficulty of items, for instance, are practically unavoidable. If a person makes a certain score in intelligence, it is somewhat dubious to say that this is his real score. Suppose that the test instead of comprising 50 items comprises a million. The latter test will probably give a more typical picture of the person's

ability. The results of the former set of items may deviate appreciably from the results of the latter. It is a question of how reliable a sample of the particular mental performance is embodied in the brief test. Or suppose that the same subjects take the same test over again. It may be found that their second scores are appreciably different from their initial scores. The crucial point, however, is whether in the second test subjects maintain approximately their initial relative standing; in other words, whether a subject who makes a good score in the first test does likewise in the second and *vice versa*.

This procedure of repeating a test and comparing initial and final scores is the accepted method for determining the reliability of a test. If, for instance, the standing of the subjects in the first test is correlated with the standing of those same subjects in the second, and this correlation is high, say upwards of .80, we consider that this test is reliable. By reliability we simply mean that the test tends to place the subjects repeatedly in the same relative position. It is rather common practice to make the test in two forms, to give each form separately to the same group of subjects and correlate the results.

**Correlating two parts of the test.** Instead of constructing the test in two forms and giving these separately to the subjects, one form only may be given, but it may be divided into two parts. If, for instance, a test comprises 100 items, we may take each subject's score on the first 50 and his score on the last 50 and correlate the two measures. If the subjects who do well in the first part of the test do well in the second, we may assume that these two groups of items are measuring practically the same thing and hence are reliable. In a time-limit test, of course, if the subjects work steadily from the beginning, obviously some of them will not complete as much of the second part as of the first, so that the two will not be comparable. In such cases, it is customary to divide the test into parts A and B, divide the time limit in two and allow the subjects the same length of time on each part. These two scores may then be directly correlated.

A similar procedure consists in evaluating separately the odd-numbered and even-numbered items in the test. This may be

facilitated by having the answers so located that the odd ones will occur in one column and the even ones in another. These may then be totaled separately to give two scores in the test and these scores then correlated.

It is thus essential after a test has been devised to investigate its reliability before putting it into practical use. In dealing with different groups of employees or applicants, we want always to apply the same mental measurement, just as in determining the dimensions of different house-lots we prefer to use a tape made of steel rather than of rubber. If an unreliable test seems to indicate vocational aptitude with one group of employees, it may utterly fail to have any prognostic value with another group.

#### SUMMARY

Mental tests such as are used for predicting vocational aptitude are devices for measuring a typical sample of mental or motor performance. Their administration may involve the examination of one individual at a time or of a group of subjects simultaneously. The latter procedure saves much time, although there is more chance for the subject to fail to follow directions and there is less chance for observation of the subject's extraneous reactions. With a small number of subjects, however, and ample floor space, the group test has most of the advantages of the individual test. The subject's response may be oral, it may be written on a blank, or it may involve some performance with implements or apparatus. The response may be entirely free or it may be constrained to various degrees. This constraint may be imposed by the wording of the question, by the location of the answer, or by necessitating the selection of the answer from two or more alternatives. Tests are given with either a time limit or a work limit. The former is generally used in group tests because it is impossible to use the latter unless the subjects can be trusted to record their own times. It is not feasible to let the examiner record the individual subjects' times in the work-limit method unless these times are relatively long, as in the "omnibus" test. The time limit must be sufficiently short so that no subjects will quite finish. Otherwise it will be impossible to differentiate between the proficiency of those who complete the

test. Certain general precautions must be observed in test procedure, such as maintaining as far as possible standard conditions, having the subject in a coöperative attitude, and having the examiner perfectly familiar with the technique so that it will go smoothly.

In selecting test material all items of a given sort may be of approximately equal difficulty with the emphasis placed on speed, or they may increase in difficulty with the emphasis on ultimate level of attainment or "power." The material is usually arranged to group together items of a given sort so as to facilitate separate scoring, but sometimes the different kinds of items are intermixed. This "omnibus" form of test may present all items of a given sort of approximately equal difficulty (cycle omnibus) or the items may increase in difficulty throughout the test (spiral omnibus). In preparing a test, alternative material of the same difficulty as the original must be provided to meet the situation if blanks reach the hands of subjects before they are tested. The test must be sensitive, i.e., give a wide range of scores. This can be accomplished by having a considerable number of items in the test and having it of appropriate difficulty, neither extremely easy nor hard, for the group taking it.

Test instructions must be kept absolutely standard and constant whenever the test is used. They should be sufficiently clear to enable the subjects to understand perfectly what is wanted. It is well to come down to the intellectual level of the lowest person in the group. Instructions usually comprise explanation, illustration, and practice. This practice is particularly important in motor tests. Printed instructions are usually preferable to oral because of their more rigidly standard character. Incentive while taking tests must be kept at a maximum in order to keep it constant. This may be done by the instructions or by utilizing various motives such as pride, coöperation, or competition.

The scoring of tests must be unequivocal so that different persons scoring the same subject's blank will obtain identical results. The blank may be arranged with a view to ease of scoring by means of stencils or by the location of the answers in convenient fashion. In obtaining a final score for a test, the question of the relative im-



portance of speed and accuracy arises. Sometimes speed is stressed and accuracy neglected, sometimes the reverse, and sometimes both are combined into a single score. This may be done purely arbitrarily or, in some instances, by considering the probability of getting correct answers by guessing. The best procedure is to correlate speed and accuracy separately with the vocational or other criterion and weight them by partial correlation technique. Scores combined in this fashion will give a more valid prediction of the criterion than those combined in any other fashion.

Before a test is put into practical use its reliability should be determined. This may be done by giving two forms of the test to the same subjects and correlating the results to see if the subjects maintain in the second instance their same relative standing. The same result may be achieved by correlating separately the first and last parts of the same form or by totaling separately the odd-numbered and even-numbered items and correlating the two. If these correlations are high, we may then use the test with impunity in employment research.

## CHAPTER VI

### THE CRITERION

#### NECESSITY

**Basis for evaluating tests.** Mental tests, like other instruments, do not always serve the purpose for which they are designed. A radio hook-up may include low loss, neutralization, and careful shielding, and yet not reach the coast. A mental test may be reliable, objective, and fool-proof, and still utterly fail to separate the sheep from the goats in the stitching room. The psychologist is no more omniscient than is the electrician. In either case it is necessary to give the instrument an actual trial and see if it does what it is supposed to do. Consequently, before psychological tests can be validly used for employment purposes they must themselves be tested by comparing, in a typical group of workers, efficiency in the tests with efficiency in the job. This implies two measures for each person on whom the tests are standardized — his test score and some figure that represents his occupational efficiency. This latter — the thing by which we are actually evaluating the tests and the thing which we wish ultimately to be able to predict — is technically termed the "criterion."

The need, however, is not merely for a criterion as such, but for as reliable and accurate a one as possible. The value of the entire project depends upon it, because it is the standard used in evaluating the tests. If the criterion is inaccurate, the tests designed to predict it will be proportionately inaccurate. If a foreman rates his men, for instance, more on the basis of physical strength than on the basis of ability in the job, it may be possible to devise tests which will correlate with his estimates, but these tests may be of little use in hiring new workers because they will predict strength rather than occupational ability. Or if the "old man" ranks his subordinates mainly on the basis of length of service rather than on the basis of actual efficiency, the tests so evaluated will predict stability rather than efficiency. Or if the production figures are

based on piece work rates that have been unscientifically and carelessly determined, the tests will not predict proficiency in that work. For the best that the tests can do is to predict the criterion by which they are evaluated. If the criterion is inadequate, the entire project is resting on shifting sands. Hence every effort should be made to get the best possible data regarding the workers' ability in the job and to handle those data in the best scientific fashion once they are obtained.

**Insuring availability of the criterion.** Before undertaking a project of this sort, it is well to make certain that the criterion will be available when needed. One should ascertain whether production records are kept in such a form that they can be utilized or whether foremen and superiors are willing to coöperate in making ratings. If the tests are given to applicants for employment rather than present employees, it is well to initiate at the very outset some procedure for following up those tested. As a matter of fact it is a good policy when possible to obtain the criterion in advance of any testing at all. If a lot of employees are tested with the understanding that subsequently the foreman will rate them and then the foreman dies, the efforts have been largely wasted. One of the committees that approached the problem of tests for aviators gave a considerable range of tests to a large number of cadets at one of the ground schools with the understanding that these men would be sent to some flying field from which subsequent record of their progress could be obtained. Many of them, however, were sent directly to France for their flying instruction so that it was impossible to obtain the criterion in their case. A few experiences of this sort impress the employment psychologist with the importance of making certain of the criterion in advance or at least of insuring its ultimate availability before undertaking any project.

As implied in the foregoing there are two outstanding types of criteria that are most frequently used. One consists of estimates made by an employee's superiors — usually foremen or inspectors in the factory and managers or supervisors in the office. The other consists of individual production figures — some objective measure of the amount of work done per unit time. In addition there are a few miscellaneous criteria that can sometimes be obtained. If

several criteria are available for a given job there then arises the further problem of reducing them to common terms and combining them into a single figure. These topics will be discussed in order.

#### ESTIMATES BY SUPERIORS

Estimates by an employee's superiors constitute one of the most frequently used criteria. Practically every member of a concern's personnel is "under" somebody else. There is some one who exercises a certain amount of supervision over him and who has some notion as to the kind of work he is doing and his value to the company. If this superior party has watched the man as closely as he ought and is willing to make estimates and is sufficiently careful in making them, they are of some value. The sort of estimates to be discussed in this connection should be distinguished, however, from the systematic rating scales to be presented in a subsequent chapter. These latter involve the separate judgment of a considerable number of traits that are not measurable by mental tests and the ratings are used in lieu of test procedure. In the present connection the estimate is usually made of only one thing, such as "efficiency in the job," and it is an estimate of something that it is hoped to predict by means of tests. The estimates that form the basis of the criterion are not usually as complicated nor are they as extensive as those involved in rating scales.

Suppose that one or more foremen<sup>1</sup> are going to make estimates of a given group of workmen. There are several ways of proceeding to the actual process of judging. The men may be simply grouped into a number of classes, they may be arranged in order from best to worst, or they may be rated systematically on a linear scale.

**Estimates by grouping.** The simplest and likewise the least reliable method of making these estimates is simply to divide the men into groups on the basis of their ability. Sometimes as few as two groups are used. The foreman is directed to divide his men into

<sup>1</sup> In the following discussion, estimates of foremen will be mentioned for the most part. While perhaps the majority of the instances encountered in actual practice involve estimates of industrial workers by their foremen, the same principles apply to office workers rated by their managers and supervisors or to executives or salesmen rated by their superiors. While foremen's estimates are used for purposes of illustration the methods described are of general application.

good and poor. He may do this by making out two lists himself or he may be given a list of all the men and be asked to check them with appropriate symbols. The difficulty with this procedure is that it assumes a dichotomy between good and poor, whereas ordinarily all degrees of ability are represented. Moreover, these data do not lend themselves to careful statistical treatment. The most that can be done is to compute the average test score made by each group and note whether the good workers exceed the poor workers in test scores, whereas it is highly desirable to use the procedure of correlation in order to be able to predict the *probability* of success in the job on the basis of the tests. Matters may be somewhat improved, if only two groups are to be used, by selecting smaller groups at the extremes of ability. If there are 100 men it may be better to select the 25 best and the 25 worst than to pick the 50 good and the 50 poor. This obviates to some extent the assumption of a dichotomy, although even within an extreme group there are doubtless marked differences in ability. The average test scores made by the two extreme groups will probably differ more than the average scores made by the two groups that comprise all the men — providing the tests are of any value at all. It may be possible to assign some arbitrary value to each group and compute a very rough correlation coefficient which will be more meaningful than if the same procedure is adopted using all the men. At any rate, if two groups only must be used, it is preferable to select them from the extremes of ability rather than to divide the entire range at the middle.

In making estimates by grouping, it is desirable, however, to have more than two groups. The foremen may be directed to divide the men into three groups — good, average, and poor; or five groups — excellent, good, average, fair, and poor. In general the more groups the better, up to a certain limit, because ability in the job is actually a continuous variable — i.e., there is a continuous gradation from worst to best — and the use of more groups gives a closer approach to such continuity. A classification into ten groups is fairly satisfactory because in correlation procedure the measures are often grouped into as few as ten classes. There is a rather delicate statistical problem involved with reference to the size of the groups. We know from the results of measurements of great

numbers of human traits that there are usually more people of average ability than of any other degree, and that as you go up or down toward the extremes the numbers decrease so that there are very few with extremely good or extremely poor ability. Probably this same thing holds true for most occupational abilities, and theoretically the estimates should comprise a large middle class with smaller classes above and below it, and still smaller classes above and below these. However, this concept is probably too complicated for the average foreman who is making the ratings and it is perhaps better either to neglect it or else to get the ratings in some quantitative form as described below and then make the proper divisions if it seems desirable. If the method of grouping is to be used at all, the safest rule is probably to use as many groups as possible (up to some reasonable limit such as twenty) and specify them by careful qualitative description.

**Estimates by ranking.** A somewhat better procedure than the foregoing is the method of ranks or the method of order of merit. It consists simply of arranging the individuals in order from best to worst. The names may be written on cards and the cards arranged in order or the names in an alphabetical list may simply be numbered. Various blanks have been devised to facilitate this process of ranking. One such involves a large card on which the names are typed alphabetically in a column in the spaces provided. At the right is another column in which the names are typed again in identical fashion. (This second column may be folded under and typed as a carbon copy of the first.) The spaces in the second column are separated by perforated lines. When the card is torn on these lines the person making the estimates has a series of miniature cards each bearing the name of a worker. He then arranges these in order on the desk with the best worker at one end and the worst worker at the other. After the order of the cards has been arranged to his satisfaction, the ranks can be transferred to the intact alphabetical list, taking the one at the best end and marking him 1 on the alphabetical list, taking the next best and marking him 2, marking the next best 3, and so on down to the worst.<sup>1</sup> This method of obtain-

<sup>1</sup> It is possible to assign two or more persons the same rank if desired. In such cases, however, for statistical reasons they must each be assigned a rank obtained

ing estimates by ranking is simple and lends itself readily to subsequent statistical treatment of the results, for it is possible to rank test scores similarly and correlate the two sets of ranks. There is one thing, however, which this method overlooks. There is nothing to indicate whether the steps between successive pairs of ranks are equal or otherwise and in handling the data the only possible procedure is to assume that they are equal. It must be assumed that the man ranked 1 is just as much superior to the man ranked 2 as is the latter to the man ranked 3. As a matter of fact this may not be the case. Suppose that the actual values of the three best persons in occupational ability or test score or anything else are represented by the numbers 75, 60, and 59. In the rank method they will be marked 1, 2, and 3, and the assumption made that the difference between 75 and 60 is the same as that between 60 and 59. This assumption, of course, entirely obscures the comparatively great superiority of the first individual. The rank method hides the light of genius under a bushel. Nevertheless, if there is a considerable number of men in the group that is ranked, this assumption of equal steps will not make such a tremendous difference, and inasmuch as the method is simple and easily administered, it is widely used.

Estimates on a linear scale are usually the most desirable. A blank is provided in which the names of the men to be rated are typed at the left of the page and each name is followed by a line of constant length. The foreman makes a check mark at some point along this line to indicate his judgment. The right end of the line may indicate highest ability and the left end lowest ability. The farther to the right the mark is placed the better man it indicates. Inasmuch as the lines are of constant length it is possible, after the ratings have been made, to convert them into figures by simply measuring the distance of each check mark from the left.

While the linear scale may be presented in the above form with a mere indication of extremes, it is better to give some notion of the intermediate steps, and, although the ultimate figures take by averaging the ranks they would have received if they did differ slightly. For instance, if they are tied for third and fourth place, they should both be ranked 3.5 and the next inferior should be called 5. If three persons are tied for fifth, sixth, and seventh places, they should all be ranked 6 and the next person below them 8.

no account of them, to provide classes or other specifications as a guide to the person making the estimates. One method is to have the blank ruled into a number of columns as a guide. There may be three of them headed "poor," "average," and "good," either of uniform width or with the average column wider than the other two. The writer has found quite successful a five-column arrangement, although it has no theoretical superiority to other arrangements that might be devised. A portion of the blank appears as follows:

NAME	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Adams	.....	.....	.....	.....	.....
Andrews	.....	.....	.....	.....	.....
Briggs	.....	.....	.....	.....	.....

It is well to arrange the width of the columns in some convenient unit. A column width of 20 millimeters gives a total width for the blank of 100 millimeters, which is a convenient maximum. The blank is accompanied by directions such as the following: "Imagine all the men you have ever known who worked at this job divided into five classes with reference to their ability in the job — a highest fifth, a next highest fifth, a middle or average fifth, a next lowest fifth, and a lowest fifth. Put a cross somewhere along the line after each man's name to indicate in which group he belongs. Moreover, if he stands high in a group, place the cross toward the right of the column, and if he stands low in that group, place the cross toward the left. In other words, the greater a man's ability in the job the farther to the right the cross is to be placed." This sort of explanation is usually intelligible to the average foreman, and it is well anyway to discuss the matter with him and bring out in an interview any misunderstandings on his part. After a blank of the above sort has been marked, it is a simple matter with a ruler to measure the distance of each mark from the left edge of the left column in some convenient unit such as millimeters or fractions of an inch. This yields a quantitative expression of the ability in



question. If the measures are to be ultimately grouped for statistical treatment into, say, 10 or 15 classes of equal size, a transparent stencil ruled with 10 or 15 columns and numbered at the top may be placed over the blank. Each check mark is noted with reference to the column of the stencil in which it falls and the number at the top of that column recorded as the criterion.

Another scheme for indicating intermediate steps between the extremes of ability is to note some descriptive adjectives along the line on which the rating is to be made. If, for instance, the criterion is to consist of an estimate of "quality of work" each line may appear thus:

.....  
 Many Careless Good Practically  
 errors quality workmanship  
 .....

These adjectives may be repeated for every line on the blank or the descriptions may occur only at the top. This procedure is used more frequently in rating scale technique where a considerable number of traits are to be rated for each man. There such descriptions are more essential for guidance as the rater is considering one trait after another. In the present situation this sort of description is perhaps not so necessary, providing sufficient explanation is given to the rater. This method of providing descriptive adjectives along the line will be discussed more at length in Chapter XII on rating scales.

One other point that applies to all the foregoing methods for securing estimates should be mentioned. An estimate of a worker before he has been at the job a sufficient length of time to reach his maximum proficiency is of little value. A man who has just recently taken up a certain occupation may be given a much lower rating by his superiors than a man who has been at the job for months, although a year later the former may be doing much more effective work than the latter. Many of us have been agreeably surprised at, or disillusioned by, the ultimate proficiency of a stenographer in contrast with our initial impression. Hence those who are making the estimates should consider whether the persons concerned have been at the job long enough to reach their ultimate

level. If the foreman is not certain about a particular man and cannot tell what his ultimate status will be, it is best to omit that man's record from statistical consideration. An experienced foreman who has trained many men will sometimes be able to estimate fairly well the ultimate status of a man who is in the earlier stages of learning his job. This practice, however, is not to be recommended and it is much better, if possible, to base statistics only on workers who have "arrived."

**Reliability of estimates.** In actual practice ratings of the sort above described often prove to be none too satisfactory. Before using them it is essential to determine their reliability, because if this is low it is apt to invalidate the entire project. By reliability is meant the extent to which the ratings agree or correlate with themselves. This is the same sort of thing discussed in Chapter V (p. 136) in connection with the reliability of test scores. We may, for instance, get the foreman to make his ratings and then at some later time, perhaps in a week or two, when he has partially forgotten the exact details of his original ratings, have him go through the process again. If his later ratings agree or correlate well with his earlier — i.e., if the same workmen are rated high in both instances and the same workmen low in both cases — his rating is more reliable than if this correlation is small. Furthermore, the ratings made by one foreman may be correlated with those made by another. If they agree closely — i.e., if each worker is rated about the same by both foremen — this indicates high reliability, but if the foremen disagree the reliability is low. In the latter case it is sometimes possible in conference to discover the reason for the discrepancies, such as personal prejudice or overemphasis of some minor aspect of the workman's performance. It may be that one foreman is stressing speed and another accuracy or that one is rating a man low because he is frequently late or because he is ugly. If these matters can be brought out in conference, it may be possible to revise the ratings somewhat and thus to get a truer indication of the actual ability in question. This will increase the correlation between the ratings or their reliability. Further investigation may be made likewise by comparing ratings with other criteria such as production (*infra*). The industrial psychologist will run into situ-

ations where the foremen concerned are simply unable to provide reliable ratings. In such cases, if other more reliable criteria are not available, it is useless to undertake the project because the ultimate value of the tests depends on the criterion by which they are evaluated.

#### PRODUCTION FIGURES

The other most commonly used criterion for evaluating the tests or other predictive measures is production. This is after all the most obvious criterion. It is the thing which the management is ultimately interested in predicting and under favorable conditions is perhaps the best indication of a man's ability in the job. In many instances it is comparatively easy to obtain because the production records are actually kept for purposes of making out the payroll. In operations such as checking, pasting, assembling, and making various parts of shoes, garments, or tires, a record of the number of pieces done per unit time is often available. Some machines such as looms carry automatic counters which record the number of operations performed. Sales records are frequently kept of men in the marketing end of industry. Even in evaluating a foreman's efficiency the production of his department may be significant.

**Workers' attitude.** There are, however, a number of problems that should be carefully noted in the particular situation before accepting the production criterion as adequate. In the first place, the attitude of the worker toward his work must be considered. His production record is not a true measure of his ability in the job unless he devotes to the job his best effort. This implies that he is industrious rather than lazy, that he is not sick or worried or engrossed in other matters, and that he has ample incentive to bring out his maximum endeavor. Some of these matters it may be impossible to ascertain at all, but frequently the foreman or supervisor will be in more or less personal touch with the men and able to supply this information. In obtaining estimates by foremen, if they are encouraged to add comments on particular men where they think it important, such points as the above will sometimes be covered. As to actual incentive, this is to a very appreciable extent

assured in the case of piece workers. These persons are paid so much per unit of work — i.e., their wage depends directly on what they do — and hence in most cases they will do their best in order to get a large pay envelope. Even in the case of piece workers, however, the foreman's judgment is by no means unnecessary, for there are instances of "stereotyping of output" in which, in spite of the possibility of more pay, men will voluntarily limit their production. This question of attitude is a much more serious problem in the case of day work in which the person is paid a flat time rate regardless of the amount of work done. It may be necessary for him merely to keep moving in order to hold the job. Often no official record at all is kept of his performance, although it is sometimes possible to collate such figures from time slips. Where the time rate is flexible it might seem that a man's rate would indirectly reflect production, but it is just about as apt to reflect his length of service, his aggressiveness in asking for a raise, the size of his family, or his consanguinity with the foreman. Sometimes records are kept with a view to determining when to promote or to raise the time rate. In such instances the incentive is partially obtained. But the production of day workers is at best a precarious criterion.

**Equivalent units of production.** Another problem that it is necessary to consider is whether the units of production used in determining the scores of different men in the same occupation are equivalent. If all the workers involved in a given study are making exactly the same piece — e.g., all making a 3½-inch fabric tire or attaching number 3 labels or typewriting form letter number 5 — the actual number of pieces done per hour by different individuals are of course satisfactory units by which to compare those individuals. If, however, one is building a 3½-inch fabric and another a 5-inch cord tire, if one is pasting small labels and another large, if one is typewriting short letters and another long letters, it is not fair to compare them in terms of tires or labels or letters. Likewise, a salesman's production may be influenced by the difficulty of the territory, by competition, by prejudice, or by whether he sells furniture or notions. In some such cases it may be possible to divide the work into smaller comparable units, such as number of lines typewritten or even number of strokes made, if appropriate

recording devices are attached to the machine. In other cases recourse may be had to the results of the time study that has been made in setting piece-work rates. If the rate has been set so that a piece that takes twice as long as another is given twice the pay, then the pay per hour is a reliable index of production. In other instances it may be better to take production as a per cent of the standard, set by time study. If, for instance, the standard is 50 units per hour and the worker does 60 units, his score is 120 per cent, whereas if he does only 40, his score is 80 per cent. If the time study has been properly done, this seems to be a fair way to get comparative production figures.

Allowance must be made for extraneous factors that influence output. If the power plant breaks down or a fuse is blown that controls certain machines, production is obviously lower, but not because of the worker's inefficiency. A loom near a doorway where there are frequent draughts has more breakages of the yarn through no fault of the operator. A belt may slip sufficiently to produce an appreciable decrease in a worker's production. In operations where persons are working together as a team — e.g., in building tires or folding tablecloths — from the standpoint of one person's production the other person constitutes an extraneous factor. The slowest worker sets the pace. This may be seen in the figures for teams of workers folding tablecloths in a laundry. (Cf. Table XI.)

TABLE XI. INDIVIDUAL PRODUCTION AS AFFECTED BY OTHER MEMBER OF THE TEAM<sup>1</sup>

TEAM	SECONDS REQUIRED TO FOLD ONE TABLECLOTH
Worker A and worker B. ....	16.3
Worker A and worker C. ....	22.0
Worker A and worker D. ....	16.1

<sup>1</sup> After Laird.

The number of seconds taken to fold the cloth by the team A and C is about 35 per cent longer than the time taken when A is teamed with either B or D. (322, 168.)

**Absences.** Another related problem concerns absence from work as affecting production. If, for instance, weekly total production is to be taken as the criterion and some men work the full week and some are absent one or more days, no fair comparison can be made. In many instances the number of hours worked per day by fellow workers varies. Consequently the safest measure seems to be production per hour. However, it is necessary to accumulate data over a considerable period, because production in any given hour may be affected by fatigue, illness, distraction, ventilation, illumination, or any one of numerous extraneous factors. If production for the first hour on Monday is the criterion selected, Smith may do poorly because he had been out late the night before; Jones may fall below his normal level because of a week-end indigestion, while Brown may have rested over Sunday and be doing his best. By the middle of the week Smith may have caught up his sleep, Jones may have recovered from the indigestion, but Brown may have been up all night with a sick friend. These discrepancies can be ironed out only by taking the average of several hundred hours' production. In some typical studies the average hourly production based on four weeks' work was taken as the criterion.

**Experience.** The amount of experience of the workers, unless considered, may vitiate the results. Those who have not been at the job long enough to reach their maximum efficiency will naturally not have a record that is typical of what their innate ability will enable them to do. All such cases should be discovered if possible and allowance made or their results excluded. They may be located frequently on the basis of the foreman's judgment. Sometimes, in a given job, it can be established by a study of the records of new workers about how long a time is required, on the average, before attaining maximum efficiency. The results of men who have not worked at the job this length of time may be excluded. Or it may be feasible to take the record of the individual in question over a considerable time and determine whether he is still improving or whether he has ceased to improve. At any rate, attention must be given to this factor of experience. If workers are used who have had time to reach their maximum efficiency and if records of piece-work production are accumulated over several weeks and reduced to pieces per hour, a satisfactory criterion will generally be obtained.

**Reliability of production figures.** In dealing with production figures, just as in dealing with ratings, effort should be made to determine the reliability of the data. If production per hour is computed for one week it may likewise be computed for some similar period and the two measures correlated. If workmen who have relatively high production per hour during one week have a similar high production during another week and *vice versa* — i.e., if the correlation between the records for the two weeks is large — we may consider this production criterion reliable. In actual practice reliable production criteria are found more often than are reliable estimates by foremen. This is probably due to the fact that the former data are more objective than the latter and do not involve personal idiosyncrasies on the part of the foreman making the estimate. More reliable estimates of a person's height could be made with an objective yardstick than by combining the subjective judgments of his acquaintances. Similarly, the more objective character of production figures gives them a distinct advantage over other criteria. If it were always possible to obtain the production record under ideal conditions, conforming to the various factors outlined above, estimates by foremen could probably be dispensed with. Unfortunately this is seldom the case.

#### MISCELLANEOUS CRITERIA

In addition to the foregoing factors — estimates of superiors and production figures — there are a number of miscellaneous things that sometimes may serve as criteria. These are not as universal as the foregoing and some of them are involved in only a limited number of occupations. In some instances they may be sufficiently reliable to supplement or even to replace the foregoing criteria.

The quality of work is one such factor. Whereas in most production records it is quantity that is noted, there are instances in which it is possible to obtain some indication of quality as well. If a considerable amount of the work fails to pass inspection the per cent that so fails may give some indication of the worker's ability. If consumers return goods on account of defective workmanship, this may often be traced to its source. If a person handles breakable materials a larger record of breakage serves to indicate a less efficient

individual. Looms sometimes carry automatic counters for recording breakages of the yarn. If the occupation is one that involves the possibility of accidents as in the case of a motorman or taxi-driver, the number of accident claims is sometimes taken as the criterion. The amount of material wasted relative to the amount used, as in cutting leather, may be determined by weighing. In these instances the emphasis is on some aspect of the quality of the work performed by the operator.<sup>1</sup>

**Amount of preliminary training.** If the concern operates a vestibule school in which the prospective workers receive preliminary training under expert supervision before being put on the actual job, it may be possible to obtain from the school records some criterion material. If the grades kept in the school are based on actual measures of skill in the work at various levels of training, these may contribute information regarding the worker's ability in the job. In some types of work the preliminary training continues until the employee has reached a certain level (as judged by his teachers) before he begins actual service or before he undertakes some special kind of job. In such instances the length of the time taken to train the man up to that point may serve as a criterion. To be sure, if the men have not all had the same opportunities for instruction or the same working conditions, a variable is thus introduced which may vitiate the figures.

**Length of service.** The actual length of time the man has been in the employ of the company is of interest. It may be that the more efficient ones remain longer inasmuch as they are successful and contented. It sometimes may be, however, that persons of high ability do not find a simple job sufficiently interesting and hence do not stick. Hence, this factor of length of service should not at the outset be taken alone as a criterion, but should be compared with actual efficiency on the job. It is often very illuminating, however, to study records of length of service or turnover with reference to occupational efficiency and with reference to mental tests.

**Advancement.** With work of an executive nature the criterion is usually more difficult to obtain. Salary is some index of an executive's ability. If a man is especially good he is generally raised



in order to keep him, while the inefficient is not raised and receives no openings elsewhere. Salary is, however, complicated by other things, such as length of service with the firm or the man's ability to sell himself to the management. Commissions are perhaps a better criterion than salary when they are involved, as they more definitely reflect output. Advancement in the firm is a related factor. In general the better man is promoted. However, account must be taken of the fact that some jobs are merely a source of supply for certain others, and being one of a considerable number of men promoted through such channels is not as indicative of good ability as being promoted in a less usual channel. Again, the responsibility which a man is given is some indication of his ability at his work.

There is one other type of criterion that is occasionally used, but which does not so often apply to actual business concerns. That criterion is membership in various organizations which require for admission some achievement in the line of work involved. Many professional organizations are of this character. An engineer, a scientist, or a professional man who is admitted to the organizations or societies in his field has probably qualified in some way. In certain studies such a thing as presence in *Who's Who* may be of significance.

These miscellaneous factors should not supplant the criteria discussed earlier. The latter are more universal and generally more valuable. In certain situations, however, some of these miscellaneous factors may be useful by way of supplement.

#### COMBINING HETEROGENEOUS CRITERIA

In obtaining the criterion it is always well to make as many approaches as possible rather than to put all of the eggs in one basket. In view of the fact that the tests ultimately developed will be no more reliable than the criterion by which they are evaluated, it is important to overlook nothing which may contribute to the reliability of the criterion. Consequently, it is well where possible to obtain production records, estimates from as many superiors as are competent to make estimates of the workers in question and any other data that may be available in the particular situation. Ac-

ording to the general principle of averages the more figures available regarding a man's ability in the job the more typical will be the average of those figures. In the ordinary industrial situation there are usually two or three superiors who can estimate most of the men in the group. It is highly desirable to have several inasmuch as one foreman may rate a particular man high or low due to prejudice and this is partially offset in averaging this estimate with those made by the other foremen. Even though every foreman cannot rate every man, due to lack of information, if most of the men are rated in common it is possible to make statistical allowances.

Given several sets of estimates made by foremen, a set of production figures and possibly some other data, the problem then arises of combining these measures into one, because in correlation procedure it is necessary to compare only two things at a time — test and job. It is obviously impossible to average production records directly in the form of pieces per hour with estimates in terms of millimeters on a linear scale or even in terms of ranks. Moreover, the linear estimates made by one foreman may not be directly comparable with those made by another. The first may rate all his men very low in ability while the second may be very lenient. A comparatively high figure assigned a man by the strict foreman may be on a par with that assigned to one of the worst men by the lenient foreman. Consequently, it is necessary to consider means for combining these heterogeneous data into a single set of values, one for each workman. Two possibilities will be discussed — when the data are in the form of rankings or order of merit and when they are in quantitative form such as pieces per hour or estimates on a linear scale.

The original data may be in the form of ranks or may be readily converted into that form. Perhaps the foremen actually make their estimates by assigning numbers 1, 2, 3, etc., to the men to indicate their relative merit. If their ratings are made on a linear scale, these figures may likewise be ranked. Production figures may be treated in the same way. Thus, if there are 20 men, there will be some number from 1 to 20 for each man in each criterion.

If every man is judged by every foreman who makes any judgments at all and if production records, if used at all, are available for

every man — in other words, if the data are complete — the procedure of combining these heterogeneous criteria is simple. It is merely necessary to average the ranks assigned a given individual to get his final ranking. A simple example<sup>1</sup> is given in Table XII.

TABLE XII

NAME	RANK IN PRODUCTION	RANK BY FIRST FOREMAN	RANK BY SECOND FOREMAN	AVERAGE RANK
Adams.....	1	2	1	1.3
Andrews....	3	1	3	2.3
Briggs.....	2	3	4	3.0
Brown.....	4	5	2	3.6
Doe.....	5	4	5	4.6

Suppose that five men are under consideration. In production Adams is the best, Briggs the next best, Andrews the third best, etc. The first foreman considers Andrews the best man, Adams second, etc. The other foreman places Adams at the top and Brown in second place. These ranks are combined in the last column by simply averaging the figures in the corresponding row. Averaging Adams's figures 1, 2, and 1 gives 1.3; averaging Andrews's figures 3, 1, and 3 gives 2.3, etc. These average ranks then give the best possible combination of the separate ranks. They may then be evaluated statistically in this form or these average ranks may themselves be ranked; e.g., Adams has the highest average rank, Andrews next, and Briggs third.

Unfortunately, in the practical situation the data are often incomplete. There will be some men who are unknown to one of the foremen and it is desirable to keep them in the data rather than discard them in order to have an adequate number of individuals in the final evaluation of the tests. It is, of course, possible where some figure is missing for a given individual to average the remaining

<sup>1</sup> The examples used throughout are usually over-simplified in the interest of clarity. Tables should not imply that a study of as few as five cases is customary or valuable.

figures. This procedure is undesirable, however, because one foreman assigns ranks perhaps from 1 to 50, while another assigns them from 1 to 43. In such a case the worst man in the one foreman's data receives a figure of 43, while the worst man in the other foreman's data receives a more severe penalty — 50. There is a statistical way out of this difficulty. It is possible to combine accurately incomplete rankings of this sort into a final set of ranks, but the technique is rather complicated. It is outlined in Appendix II.

**Combined quantitative estimates.** When the data consist of estimates on a linear scale or production figures or are in some other quantitative form rather than in the form of rankings in order of merit, a different procedure is necessary for combining the heterogeneous criteria. If the data are complete and consist entirely of the same sort of thing, such as estimates on the same type of linear scale, it is possible to average the figures for each workman with some validity. Even under these circumstances, however, error may be introduced by the fact that different judges use different standards, some being more lenient and some more severe. If, however, the estimates are not complete there is much opportunity for unfairness. The omission of the estimate of a man by a lenient foreman who has estimated all his fellows will be a distinct penalty inasmuch as his average will have to be based only on the more strict estimates, while his fellows have a lenient estimate to raise their average. Moreover, when dissimilar criteria in entirely different units are used, such as linear distances on a scale and pieces per hour produced, it is impossible to average such data in the original form just as it is impossible to average pounds and kilograms without converting them to a common basis.

One of the best methods for making such criteria comparable is to convert the figures of a given sort into terms that are relative to all the figures of that sort; e.g., to convert each estimate made by a given foreman into terms of all the estimates made by that foreman. If the first foreman's estimates average 70 and the second's average 50, an estimate, say, of 40 by the first would indicate a poorer man than an estimate of 40 by the second, because the first is setting a much more lenient standard and the mark of 40 is much lower relative to that standard. Hence one feature of importance in mak-

ing such estimates comparable is to determine how much each one is above or below the average made by that foreman; i.e., how it compares with the standard he sets. It might seem then that it is merely necessary to express the results as deviations from the average concerned. One set of deviations, however, may be in terms of linear units and another in terms of production. Moreover, even with foremen's estimates one man may bunch his estimates together rather closely, whereas another may scatter them over a considerable range. If two foremen give the same average rating of 60, but one places every man between 30 and 90, and the other rates some as low as 10 and as high as 110, an individual who is rated 30 by the first is doubtless inferior to one rated 30 by the second. Hence it is necessary not only to consider how much a given rating deviates above or below the average rating made by that foreman, but also to consider that deviation relative to the general scatter or variability of that foreman's ratings. In determining such scatter or variability it is not sufficient merely to consider the range — i.e., the highest and lowest figures given by a foreman; but it is necessary rather to ascertain how much his ratings in general deviate from the average. A simple procedure involves computing the deviations and averaging them. A brief illustration is given in the first part of Table XIII.

Suppose the 5 men whose names appear in the first column receive the ratings by a foreman indicated in the second column. Adams is rated 30; Andrews 50, etc. The average rating is 60. Adams's rating of 30 is 30 less than this 60, or his deviation is  $-30$ . Andrews's rating of 50 deviates from the average to the extent of  $-10$ . Similarly, Briggs is 30 above the average; Brown does not deviate at all, hence his deviation is 0; and Doe is  $+10$ . These deviations appear in the third column. If we now neglect the signs and average these deviations, we have 16 as the *average deviation*. As its name implies, it indicates how much on the average the different scores deviate from the average score. If we turn in the second part of the table to the ratings of these same workmen by a second foreman, we see that they likewise have an average of 60, but it is obvious that they scatter more. Computing the average deviation in the same way, we obtain 36, just twice as much as in

the first case. This gives a quantitative expression for the greater scatter in the second case.

TABLE XIII. ILLUSTRATING AVERAGE DEVIATION AND STANDARD DEVIATION

WORKMAN	RATING BY FIRST FOREMAN	DEVIATION FROM AVERAGE	DEVIATION SQUARED	RATING BY SECOND FOREMAN	DEVIATION FROM AVERAGE	DEVIATION SQUARED
Adams...	30	-30	900	10	-50	2500
Andrews.	50	-10	100	50	-10	100
Briggs...	90	30	900	30	-30	900
Brown...	60	0	0	120	60	3600
Doe.....	70	10	100	90	30	900
Total...	300	80	2000	300	180	8000
Average .	60	16	400	60	36	1600
Standard deviation			20			40

In most statistical work, where interest is in the variability or scatter of a series of measures, it is customary to use the "standard deviation" rather than the average deviation. This is obtained as follows (cf. the last column in each part of Table XIII): Instead of throwing away the signs as above suggested, the deviations are squared, thus automatically making all signs plus. For instance, Adams's deviation of - 30 squared gives 900; Andrews's deviation of - 10 squared gives 100, etc. These squares are then totaled giving 2000, averaged to get 400 and the square root taken to give 20 for the *standard deviation*. This is the same sort of measure as the average deviation, but it has the advantage that it fits into the mathematical theory of probability and actually occurs in the equation for the normal frequency curve (*infra*). Its computation by the above method becomes tedious when a large number of individuals are involved, but statistical short cuts are available. (498, 603.)

Returning now to the significance of a rating of 30 assigned by the first foreman compared with a rating of 30 by the second: Adams is

rated 30 by the first foreman and Briggs is rated 30 by the second. What is their relative ability? In both instances 30 indicates a workman below the average and in both instances he is 30 points below the average; i.e., his deviation is  $-30$ . But it is obvious that Adams stands lower in the estimation of the first foreman than does Briggs in the estimation of the second, because the first foreman's ratings do not scatter as much and hence  $-30$  represents a relatively larger deviation. The best way to express this fact is to take the ratio of the deviation to the standard deviation. If the rating of  $-30$  made by the first foreman is divided by 20, his standard deviation, we have  $-1.5$ . This means that Adams is rated below the average by an amount equal to 1.5 of the standard deviation. Taking Briggs's rating of 30 or deviation of  $-30$  given by the second foreman and dividing it by that foreman's standard deviation of 40, we have  $-.75$ . This means that Briggs is rated below the average by an amount equal to .75 of the standard deviation. These figures of  $-1.5$  and  $-.75$  show the relative significance of the same rating of 30 when made by two foremen who have a different variability in their ratings.

When estimates are converted into the above form of deviation divided by standard deviation, they are directly comparable, if it is assumed that the ratings made by each foreman follow a normal frequency curve. This is the type of curve obtained in most cases where a large number of persons have been measured in some mental or physical characteristic. The majority of persons score near the average and the farther we depart from the average in either direction the fewer individuals we find. The method of obtaining the curve can be made clear by an actual example. Scores made by about 2400 college freshmen in an intelligence test are as follows:

1 student scores between 900 and 909 points						
3 students score between 190 and 199 points						
15	"	"	"	180	"	189
35	"	"	"	170	"	179
53	"	"	"	160	"	169
103	"	"	"	150	"	159
150	"	"	"	140	"	149
181	"	"	"	130	"	139
156	"	"	"	120	"	129

282 students score between 110 and 119 points							
276	"	"	"	100	"	109	"
299	"	"	"	90	"	99	"
276	"	"	"	80	"	89	"
211	"	"	"	70	"	79	"
189	"	"	"	60	"	69	"
77	"	"	"	50	"	59	"
38	"	"	"	40	"	49	"
14	"	"	"	30	"	39	"
4	"	"	"	20	"	29	"
4	"	"	"	10	"	19	"

To plot these data in the form of a frequency curve (see Figure 1) we lay off these scores along the base line and for each erect a per-

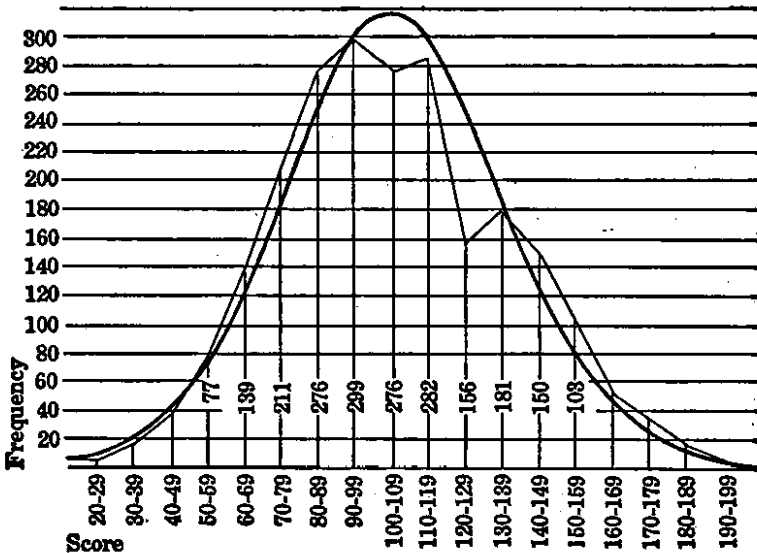


FIG. 1. NORMAL FREQUENCY CURVE

pendicular at the point corresponding to that score, the height of which is proportional to the number of students making that score. Beginning at the extreme right of the curve for the score 200-209 we erect a perpendicular of height 1; for the score 190-199 we erect a perpendicular of height 3 (these scarcely show in the figure); for



the score 180-189 we erect a perpendicular of height 15; for the score 170-179 a perpendicular of height 35, etc. The line joining the tops of these perpendiculars constitutes the frequency curve. In actual practice the perpendiculars are usually omitted. This frequency curve shows the trend above mentioned, that there is a prevalence of mediocrity and that the number decreases as we go above or below the average. Hence this curve is approximately a normal frequency curve. Most curves show minor irregularities like the present one but the general trend is usually obvious. The ideal type of curve is smooth like the heavy one indicated in the figure. If we know the average and the standard deviation of a set of data it is possible to derive the equation of the curve and plot a smooth one such as that shown. Foreman's ratings and production figures yield approximately this same sort of normal frequency curve. A given set of figures may be plotted to note whether they approximately conform and if they do so the theory of the normal frequency curve may be applied to the data.

The equation of the normal frequency curve is known and it is a function of the standard deviation; i.e., the standard deviation occurs in the equation of the curve. The properties of the curve are such that it is possible to tell what proportion of the individuals fall between the average score or rating and any other score, providing this latter is converted into terms of standard deviation. Let us recur to the preceding example of the two foremen each furnishing an average rating of 60, but the first having a standard deviation (expressed by  $\sigma$ ) of 20 and the second having a standard deviation ( $\sigma$ ) of 40. Using these figures, we may plot the two normal frequency curves shown in Figure 2. The ratings occur along the base line and the height of the curve at any point represents the proportion of the men receiving the corresponding rating. It is to be noted that both the curves have the same general shape, but the upper one is much steeper than the lower. This corresponds to the fact that the first foreman has a smaller variability in his ratings — i.e., he bunches them together. Let us now express the scores along the base line in terms of standard deviation ( $\sigma$ ). In the case of the first foreman a rating of 80 represents a deviation of + 20, and, inasmuch as the standard deviation is likewise 20, this rating repre-

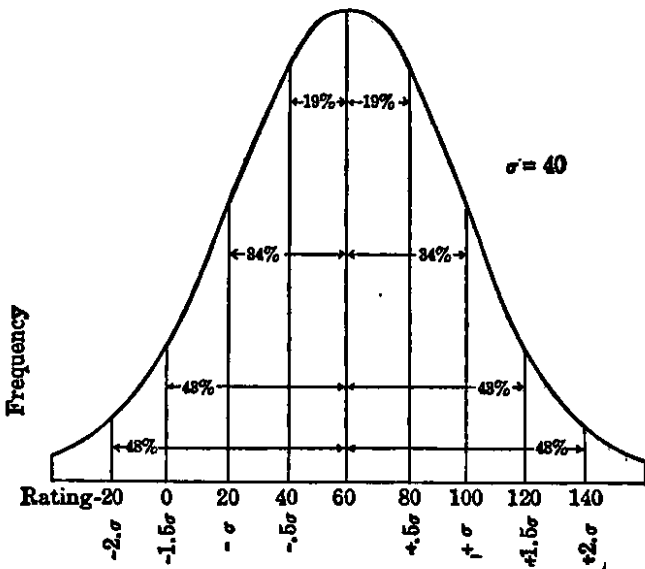
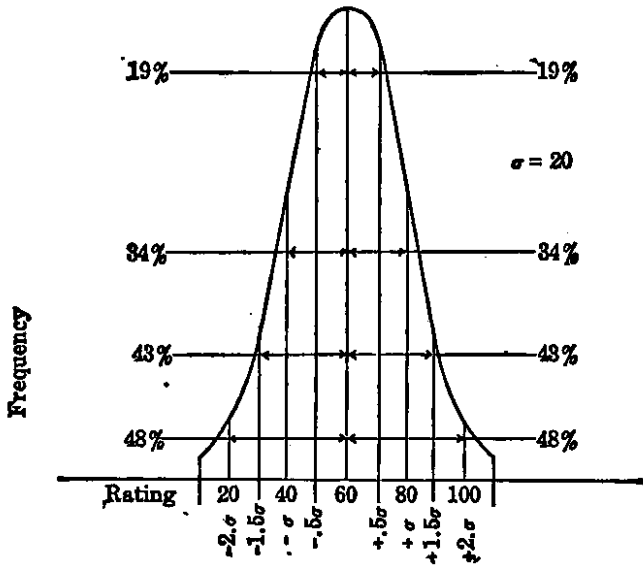


FIG. 2.

sents a rating that is above the average by an amount equal to the standard deviation or it may be expressed as  $+\sigma$ . In the case of the second foreman, a rating of 100 represents a deviation of 40, which is just equal to the standard deviation and so may be expressed similarly as  $+\sigma$ . In the case of the first foreman, a rating of 100 has a deviation of 40, which is twice the standard deviation and may be expressed as  $+2\sigma$ . Similarly with the second foreman a rating of 140 is  $+2\sigma$ . The same procedure holds with the ratings less than the average. The first foreman's rating of 40 deviates from the average in the negative direction by an amount equal to the standard deviation, and hence may be expressed as  $-\sigma$ , while his rating of 20 is less than the average by an amount equal to twice the standard deviation ( $-2\sigma$ ).

The equation of the normal frequency curve now tells us (by the use of calculus) that between the perpendicular erected at the average and that erected at  $+\sigma$  is found 34 per cent of the area of the curve (see figure). This means that 34 per cent of the foreman's ratings fall between these limits. Hence we may say that 34 per cent of the ratings made by the first foreman fall between 60 and 80, while 34 per cent of the second foreman's ratings fall between 60 and 100. Similarly, we know that between a perpendicular erected at the average and one erected at  $+2\sigma$  is found 48 per cent of the area of the curve. This means that 48 per cent of the ratings made by the first foreman fall between 60 and 100, while 48 per cent of those made by the second foreman fall between 60 and 140. A workman who is rated 100 by the first foreman is actually exceeded in ability by only 2 per cent of all the men that foreman has rated, while a man rated 140 by the second foreman is likewise exceeded by only 2 per cent of the men rated. Hence those two workmen have the same ability in the estimation of the two foremen. Likewise a man rated as 80 by the first foreman is the same as a man rated 100 by the second because each of them is exceeded by only 16 per cent of the group (i.e., 50 per cent minus 34 per cent). Similar reasoning would apply to men rated 90 by the first and 120 by the second, for both represent  $+1.5\sigma$  and are exceeded by only 7 per cent of the group. The same reasoning applies to the ratings below the average. A rating of 40 in the first instance is equivalent to one

of 20 in the second, because both are  $-\sigma$  and indicate that the individual is superior to only 16 per cent of the group.

It is now obvious that two ratings that deviate from the average by the same proportion of the standard deviation are equivalent. They are both located at the same point on a normal frequency curve and indicate the same relative standing in the estimation of the foremen concerned. Thus, if we convert original ratings or measures into these terms they are directly comparable, because being a certain fraction of the standard deviation above the average in the estimation of one foreman is the same thing as being that same fraction of standard deviation above the average in the estimation of the other foreman. In other words, we have reduced the measures to common terms, namely, their location on a normal frequency curve, and all normal curves have the same characteristics.

Exactly the same procedure may be followed with production figures or any other criterion that may be put in quantitative form. A person who is twice the standard deviation above the average in production is exceeded by only 2 per cent of the individuals, just as a person who is twice the standard deviation above the average in a foreman's opinion is exceeded by only two per cent of the individuals. This technique, of course, assumes that the data follow a normal frequency curve. This assumption would be absurd with five cases, as in the above example, which is simplified merely for illustrative purposes. However, if a reasonable number of individuals are involved, the assumption may be made for practical purposes and will make the measures more nearly comparable than if they are treated in some arbitrary fashion. When measures have been converted into this form, it is then possible to average the different measures for a given workman.

Recurring to the previous example, we may take the deviations of the first foreman's ratings and divide each by the standard deviation, do the same for the other foreman's ratings and then average the two converted measures for each workman. The results are given in Table XIV. Each of the deviations for the first foreman is divided by 20, giving  $-1.5$  for Adams,  $-.5$  for Andrews, etc. The deviations of the second foreman are each divided by 40, giving  $-1.25$  for Adams, etc. The two converted measures for each work-

man are comparable and can be validly averaged. Adams's two converted ratings of  $-1.5$  and  $-1.25$  are averaged to give  $-1.37$  in the last column. Andrews's two values of  $-.5$  and  $-.25$  give an average of  $-.37$ . One of Briggs's converted ratings is  $+1.5$  and the other is  $-.75$ . The algebraic sum of these is  $+.75$  and the average  $+.37$ . These averages that occur in the last column give the final figure that is to be taken as the criterion for each workman. Production figures or other quantitative criteria can likewise be converted into this form and then all the available figures for each workman averaged to get his final criterion. If a few ratings are missing here and there, the remaining ones can be validly averaged for the man concerned because the omission of a rating, for instance, by a lenient foreman will not unduly penalize that man. The lenient foreman's ratings are all converted to these relative terms so that they are not in their final form any more lenient than the ratings of another foreman.

TABLE XIV. CONVERSION OF THE RATINGS OF TABLE XIII INTO TERMS OF STANDARD DEVIATION

WORKMAN	RATING BY FIRST FOREMAN	DEVIATION FROM AVERAGE	DEVIATION DIVIDED BY STANDARD DEVIATION (20)	RATING BY SECOND FOREMAN	DEVIATION FROM AVERAGE	DEVIATION DIVIDED BY STANDARD DEVIATION (40)	AVERAGE OF THE CONVERTED MEASURES
Adams.....	30	-30	-1.5	10	-50	-1.25	-1.37
Andrews.....	50	-10	-.5	50	-10	-.25	-.37
Briggs.....	90	+30	+1.5	30	-30	-.75	+.37
Brown.....	60	0	0	120	+60	+1.50	+.75
Doe.....	70	+10	+.5	90	+30	+.75	+.62
Average.....	60			60			

The foregoing technique makes it possible to obtain a miscellaneous set of criteria and combine them into a single measure for each individual. They are all reduced to common terms, and even the omission of some estimates or other criteria in the case of some

individuals will not appreciably vitiate the results. This combined criterion can then be used in the project of developing tests to predict capacity for a given occupation.

#### SUMMARY

The criterion is an index of occupational proficiency which is used in evaluating the tests designed to predict that proficiency. It should be derived as carefully as possible because the tests are devised with a specific view to correlating with the criterion and if the criterion itself is inaccurate the entire project is likewise. In order to avoid wasted effort, it is advisable to insure at the outset the ultimate availability of adequate criterion data.

Estimates as to ability in the job by an employee's superiors constitute a frequently used criterion. This estimate may involve dividing the workers into two groups, but this procedure has little statistical value. It is better to select two groups at the extremes of ability, but is better still to have a considerable number of groups so that the criterion may more nearly represent a continuous gradation from best to worst. The estimate may also be made by ranking, i.e., arranging the workmen in order from best to worst. This method, however, assumes that the differences between adjacent ranks are equal and often obscures an outstanding instance of superiority or inferiority. Estimates on a linear scale are probably more desirable than the preceding. The name of each workman is followed by a line of constant length. The rater makes a check mark at some point along this line to indicate his judgment. The farther from the left the mark is placed, the greater the ability indicated and a measurement of this distance gives the criterion. As a guide to the rater the blank may be divided into columns or have descriptive adjectives or phrases at various positions along the line. The reliability of such estimates should be ascertained by correlating two sets of ratings made by the same foreman on different occasions or by correlating the ratings made by one foreman with those made by another.

Another, and perhaps, in the long run, better, criterion consists of production figures. After all, production is the thing which it is ultimately desired to predict. In many industrial operations

records of actual production per unit time are readily available. There are, however, several factors that may operate to invalidate the production figures unless taken into account. One is the worker's attitude. If he has not done his best at his work, the production is not a true measure of his proficiency. This attitude is assured to a much greater extent in the case of piece workers than in the case of day workers. In evaluating such figures it is necessary to ascertain if the units of production used for different workers are equivalent. If the men are not all making the same product, it is sometimes possible to adjust the figures on the basis of time-study records so that the units can be made equivalent and hence every one measured by the same standard. Allowance must be made for extraneous factors, such as temporary technical faults in the machinery and the pace-setting for the operation by other workers collaborating with the man under consideration in a "team." Absences from work influence the weekly production, so that it is better to reduce the figures to production per hour, but it is desirable to average a great many hours to obviate chance errors due to health or fatigue. With the production figures as with the estimates the reliability should be determined by correlating production over one period with that over another.

In addition to the foregoing there are miscellaneous criteria that are available in some instances. One of these is the quality of work as indicated by amount passing inspection, breakage, accident claims, or amount of material wasted. Another of these criteria is the amount of preliminary training given the applicant in a vestibule school or elsewhere before being put on regular work or before being advanced to a particular kind of more complex job. Length of service and advancement in the firm give some notion regarding the proficiency of executives and others where it is difficult to obtain other indications of proficiency.

After the various criteria have been obtained, the problem arises of combining them into a single figure for each individual. If the data are in the form of ranks and are complete — i.e., if there is a rank assigned to each man in each criterion — it is a simple process to average the ranks assigned a given individual to get his combined rank. If the rank data are incomplete the procedure of combi-

nation is complicated. If the criteria are in quantitative form, such as estimates on a linear scale or production figures, the best procedure is to convert, for example, estimates by a given foreman into terms of the other estimates made by that same foreman. If the foreman's estimates are averaged and the standard deviation computed to indicate their variability or scatter, and then a given estimate is converted into terms of its deviation from the average divided by the standard deviation, this estimate is then located definitely on a normal frequency curve for that foreman's estimates. It is necessary to assume that the foreman's estimates conform approximately to such a normal curve. If estimates of other foremen and production figures are converted into similar terms, they are all comparable because they are all located on normal frequency curves and the properties of such curves are universal. The measures are thus in common terms and can be validly averaged into a single figure. This combined criterion figure for each workman may then be used in the project of developing tests to predict occupational proficiency.



## CHAPTER VII

### THE SUBJECTS USED IN EVALUATING TESTS

#### GENERAL CONSIDERATIONS

IN devising tests for vocational capacity or proficiency it is necessary to standardize them on a typical group of individuals by comparing test scores with the criterion. These persons used in evaluating the tests are technically termed "subjects." The problem naturally arises as to who shall be used as subjects for the project and there are several important considerations involved in their selection. In the first place, the subjects used in standardizing the tests should be typical of the applicants for employment to whom the tests are ultimately to be given for practical purposes of prediction. Test standards obtained on college students, for instance, would be unsatisfactory for hiring unskilled laborers. Secondly, the incentive or attitude of the subjects should be similar to that involved in the ultimate employment situation. If the test is evaluated on men who do not do their best, the standards will be too low for valid prediction of the capacity of men who exert maximum effort when being tested with reference to employment. In the third place, the previous experience or training of the subjects should be taken into consideration. It is possible that some of the tests will measure factors that are influenced by a man's industrial experience, although they purport to measure innate capacity. In the fourth place, the availability of the criterion must be considered. If it is not forthcoming at the outset, the entire project will be delayed or perhaps vitiated altogether. A further problem arises if a limited group of men of a given sort is to be tested, i.e., if some selection of subjects is involved. One must determine how many subjects are necessary and how they are to be selected. Finally, there are several miscellaneous factors to be considered such as age, sex, sensory defects, and literacy of the subjects.

APPLICANTS FOR EMPLOYMENT AS SUBJECTS

**Typical of subsequent applicants.** There are two possible fields from which to select the subjects. Applicants for employment may be tested as they are hired for the job in question, or employees who are actually working at that job may be used. The former would seem at first glance to be the logical subjects for the project. They are quite typical of the men to whom the tests are ultimately to be given for employment purposes — in fact they represent exactly the same class. The only difference is that the applicants tested at the outset are an experimental group and their employment does not depend on their test accomplishment, whereas with the later applicants the test may actually determine whether or not they are hired. The applicants in the first group, however, need not know this fact. The general situation in the employment office where the tests are given will be quite similar with the experimental group and with the subsequent applicants.

**Incentive.** In testing applicants the attitude of the subjects is likewise of a desirable sort. Inasmuch as they feel that their employment depends to some extent upon their efficiency in the test, they will doubtless have a maximum incentive. This is highly desirable because we have previously seen that the only way to keep incentive constant is to keep it maximum. If the experimental applicants and the subsequent applicants both have this maximum incentive their results are directly comparable.

**Previous training.** With applicants the uniformity of their previous training with reference to the job in question is apt to be greater than in the case of employees. Sometimes, of course, it is desired to measure actual trade proficiency rather than potential capacity. Usually, however, we are dealing with occupations for which previous preparation is of little value, for the job involves a new set of specialized operations which the man must be taught, such as building a tire. The psychologist's interest is not in what training a man has had, but in the innate capacities — attention, motor coordination, or reaction time — that will enable him to make a good tire-builder after he has had requisite instruction at the plant concerned. Whereas the employees may have had some of the abilities that are measured by tests modified somewhat by their

work on the job in question, the applicants are homogeneous in this respect, as they have had no experience of this sort.

**Delay of criterion.** The foregoing facts indicate the desirability of having applicants at the employment office as subjects on whom to standardize the tests. There is one very serious drawback, however; namely, the criterion will not be available for a considerable time. If the men are tested when hired, it is necessary to wait weeks or months to determine whether they are good or poor in the job. This obviously delays the entire program of evaluating the tests by comparison with the criterion. In the majority of cases this one disadvantage outweighs the advantages of using applicants in the evaluation of occupational tests.

#### EMPLOYEES AS SUBJECTS

**Incentive.** Employees are more frequently used in such projects. The question of attitude and incentive is, of course, more serious than with applicants. But in the chapter on test technique devices were suggested for controlling this factor. The wording of the test instructions may be such as to impress upon the subjects the importance of doing their best. Their coöperation may be enlisted or they may be effectively motivated by appeals to pride or by competition.

**Previous training.** Inasmuch as the employees will have behind them different lengths of service on the job in question, the problem of what the test measures is more acute. Does a particular test measure a man's innate capacity which made him potentially a good or poor workman at a given job or does it measure traits that have been modified by his training in that job? The general theory of employment tests is based primarily on the first of these alternatives, because applicants come to the office with no experience in the job in question and it is desired to determine whether they have the innate capacity that will enable them to succeed. Hence it is important to determine whether the test measures this sort of thing. The simplest way to obtain this information is to correlate the tests in question with the length of service in the job. If, for instance, the men who have been for a long time employed as tire-builders do better on a certain motor coördination test than do those more

recently hired, this indicates to some extent that the function measured by the test is influenced by experience at that particular job. The logical thing is then to discard that particular test. There is also another alternative that is comparatively satisfactory. If all the employees used as subjects have had a very considerable experience in the particular job so that they have very definitely reached their maximum efficiency in the job, the presumption is that they have likewise improved in the test function as much as they ever will. Consequently, their results are comparable with one another rather than being vitiated by the fact that one man has had more training by which to profit than has another. Even then, however, this group would score more highly than would a group of applicants and some discount would be necessary in setting a test standard for the latter. It is desirable then at the outset to select tests which measure innate capacity. The psychologist is familiar with tests in which the subjects after relatively little practice reach their maximum efficiency. Practice in the test itself would surely improve the function in question as much as practice in the job, and, if it can be demonstrated that after a certain amount of practice on the test itself, subjects do not improve appreciably, and if that amount of practice is given to those taking the tests, it is reasonable to consider the results valid regardless of length of service. The latter is then of concern only in evaluating the criterion. At any rate, it is necessary to consider carefully whether the tests measure innate or acquired capacity and make the appropriate adjustments as above suggested.

**Availability of criterion.** With reference to the criterion it is, of course, obvious that this can be almost immediately obtained, when dealing with employees. This makes it possible to start at once the procedure of comparing test score and criterion. Such methods as are developed from this procedure can then be put into effect at an early date. If a psychologist is employed to establish such methods of selecting employees, it is probably better to sacrifice the slightly greater reliability obtained by testing applicants in the interest of avoiding delay and of getting something of practical value started as soon as possible.

## SAMPLING OF SUBJECTS

**Number desirable.** Assuming, then, that a group of employees in a given job are to be used for evaluating the tests, there arises the problem of whom to test. The first consideration is how many to include. In occupations which involve a relatively small number of workers, say not over fifty, it is desirable to test them all. If, however, there are several hundred, it is often advisable to take a sampling, i.e., to select some who will be typical and base the results on them rather than to go laboriously through the entire group. From the standpoint of the management the fewer employees taken away from their work the better, provided equally valid results can be obtained. No arbitrary minimum number can be laid down. From the statistical standpoint there is, of course, no danger of getting too many. As the number of subjects increases, the correlation between test and criterion approaches more nearly to the true correlation that would be obtained with an unlimited number. If the correlation is obtained with a small group and the procedure repeated with another group, the second result is liable to differ quite considerably from the first. A few anomalous individuals in one or the other instance may be sufficient to throw the results out considerably. With larger groups this is less apt to happen because the anomalous cases will be to a greater extent absorbed by the law of averages. A point exists in a given project, however, at which the addition of further numbers of individuals does not improve matters to any considerable extent. It is possible to determine this empirically in correlating a specific test score with the criterion by computing the correlation with say fifty individuals, then with sixty, then with seventy, etc., until the addition of ten more makes little difference in the correlation. Often, however, the psychologist has to state in advance how many men he will need and stick to his statement. In actual practice one occasionally sees reports of research based on as few as ten individuals. This is probably too small a number to be very valuable. Thirty or forty sometimes prove fairly satisfactory, but it is probably desirable to get at least fifty and preferably more. No definite minimum number can be specified, but it is doubtless better to err in the direction of too many rather than too few subjects.

**Sampling by foremen.** If a sampling is to be made there are various possible methods. The foreman may simply be asked to send over fifty of his men on some convenient schedule. This lets the foreman make the selection and it is a dubious procedure. On the one hand, he may be governed in his selection largely by convenience and send men who can be most easily spared at a given time. This not infrequently would mean that the poorer men were chosen. On the other hand, he is liable to go to the opposite extreme, wishing his department to make a good showing in the tests, and send only his best men. Either of these procedures is unsatisfactory because what is desired for statistical purposes is the entire range of ability. There is need for the good, the average, and the poor in order to determine how the different degrees of occupational ability compare with ability in the tests. A correlation coefficient computed when the range of one variable is restricted — as in the present case with only the good workers instead of all degrees of ability — is smaller than that coefficient will be if the entire range is included. Unless elaborate formulæ are used for correcting this coefficient (272, 225), it is apt to be misleading.

**Sampling by the psychologist.** It is far better for the experimenter to make his own sampling. He can then insure that the entire range of occupational ability is represented in the data. If he is making the selection in advance of the criterion, he had best secure a list of all the men from whom the selection is to be made and then take them alphabetically or else write their names on cards, shuffle the cards, and pick at random the desired number. This chance procedure will in the long run insure a "normal" distribution of ability. If the criterion is available before the testing is undertaken, the foregoing procedure may still be followed or else the selection may be made with a view to getting a normal distribution. The psychologist can select a rather large number of employees near the average score in the criterion, smaller numbers above and below this average group, and still fewer numbers as the extremes are approached. He will have to be governed somewhat by the actual appearance of the data in determining just which ones to select and will have to exercise considerable judgment, but if familiar with normal frequency curves he will have little diffi-

culty in selecting a group whose criteria distribute in normal fashion.

After a sampling has been made, the men on the final list can be examined at times that are mutually most convenient for all concerned. The scheduling of tests will, of course, depend on the local circumstances. In this connection, however, it is well to provide an alternate for each man or at least an alternate who may be substituted for any one of several because of contingencies that may arise. A few of the men on the original list may leave before their turn comes or they may be put on the night shift for a few weeks so that it will be necessary to test instead some alternate of approximately the same occupational ability.

#### MISCELLANEOUS FACTORS

**Sex.** There are a few other factors that should be taken into consideration with reference to the subjects involved in a research of the above sort. Workers may be of either sex. In the majority of cases all the workers on a given job will be of the same sex and of course no problem arises. If the tests are standardized on men and used for hiring men, the sex problem is not germane. There are instances, however, in which both men and women are employed in the same job. If there is a sufficient number of each sex it is probably preferable to evaluate them separately just as if they were two separate jobs from start to finish. From the standpoint of the criterion there is a danger that foremen or forewomen will use somewhat different subjective standards in rating subordinates of their own sex or of the opposite sex. There is also a possibility that the time-study results will be influenced by various notions regarding the relative competence or industrial value of the sexes. From the standpoint of the tests themselves there is likewise the possibility that the scores will be influenced by mental sex differences. Psychologists are at present uncertain regarding these. Experimental evidence shows that sex differences in actual ability of the sort usually measured by mental tests are slight. (248.) There are recent indications, however, that while differences do not exist in the field of ability they do exist in the realm of interest, attitude, and emotion. (328.) Most of the experimental data on sex dif-

ferences are based on pupils in college or school with whom nothing particular was at stake in taking the tests. It is possible that under the more stimulating conditions of examination in an employment office differences of this more subtle character may influence the results. At any rate, there is nothing to be lost by a separate evaluation of the sexes and there is a possibility that error may be thus avoided.

**Age.** Another factor to be considered is the age of the subjects. If there are certain mental capacities that do not reach their maximum until relatively late in life, or if there are others that begin to decline relatively early, these may make a difference in test results. Consequently, in testing persons in their teens or well along in middle life, it is desirable to consider whether the test under consideration appreciably reflects the age factor.

Many tests have been standardized on persons of different ages so that it is possible to plot curves showing how proficiency in the test varies with the subject's age. A few typical results are shown in Figure 3. The figures along the base line represent age. To make the curves comparable the average score attained by subjects of a given age is reduced to a per cent of the maximum average score made by any age (in most of the available instances this maximum score is for age 18). These per cents are plotted by locating a point directly above the given age on the base line at a distance from the base line proportional to the per cent in question. For instance, at age 6 the score in tapping is about 61 per cent of the maximum attained at age 19; at age 7 this per cent has risen to about 66. The curve for logical memory begins at age 8 with 70 per cent of the maximum score; at age 9 this has increased to 79 per cent. The heavy straight line indicates the per cent that a given age is of 18. This heavy line shows the progress to be expected if development in the various capacities measured is directly proportional to age. The data are taken from various sources involving different groups of persons and different numbers in the groups and include averages for each age based on boys and girls combined. While other experiments and different treatment of data might show somewhat different results, the present curves are probably sufficiently typical to indicate certain trends.



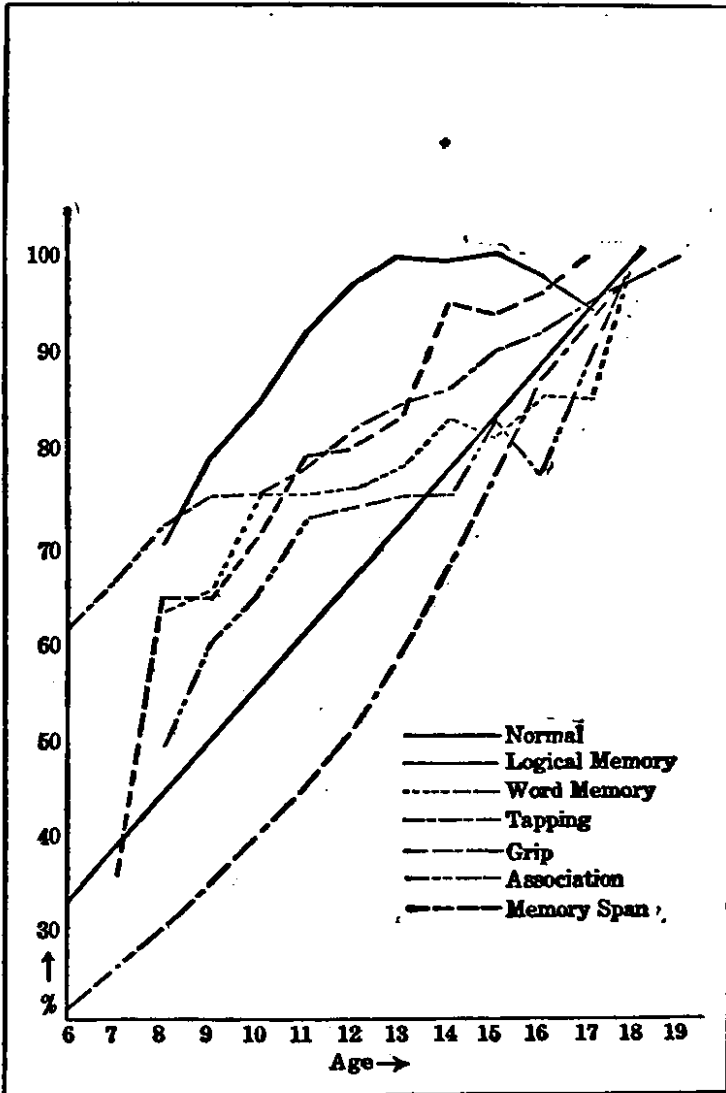


FIG. 3. EFFECT OF AGE ON TEST PERFORMANCE  
(Data from Fyle, Gilbert, Smedley, Hollingworth, Whipple)

## THE SUBJECTS USED IN EVALUATING TESTS 181

The curves all show obviously a general rise with advancing age. They do not all manifest, however, the same consistent rate of rise. Muscular strength as far as indicated by the grip increases consistently from year to year. Rate of tapping does likewise, although the curve is not as steep, and in the later teens it approaches more closely to maximum proficiency. This suggests that immature workers are more suited to work requiring rapid muscular movement than to work requiring muscular strength. Different kinds of memory show varying rates of increase with age. The simple rote memory measured by memory span (cf. example 23, Chapter IV) rises steadily to its maximum at age 17. Logical memory, however — i.e., memory for ideas in a story that is read — reaches its maximum at about age 13 and remains practically constant thereafter. Memory for disconnected words shows a period of little progress in the early teens followed by a subsequent jump. Consequently, in giving memory tests to young employees or applicants some types of test will presumably be vitiated by the age of the subjects unless allowance is made, while with other types this will not be the case. A free association test likewise has a period of little progress followed by a subsequent rise.

These curves are typical of the differences obtained with other kinds of tests and indicate, with subjects who are not adults, the desirability of taking account of age. It is obviously possible for some applicants to obtain a score which from the standpoint of predicting their ultimate proficiency in the job will be unfair, because at the time of testing their immaturity is conducive to a lower score. In such a project it is desirable to study the scores made by persons of various ages in a given test and determine at what age improvement due to maturity ceases.

Turning to the other end of the age scale there is a similar possibility that a person will make a lower score in the test because of his advanced age. In one instance a number of tests were given to subjects of varying ages up to 40. The average scores made by subjects of each age group appear in Table XV. Each group involves about 100 individuals. With the first three tests there is little difference in the average scores made by subjects of different ages. In the case of the substitution test (similar to example 14,

TABLE XV. AVERAGE TEST SCORES FOR GROUPS OF DIFFERENT AGES<sup>1</sup>

TEST	BELOW 20	20 TO 25	25 TO 30	30 TO 35	35 TO 40
Completion....	31	32	32	32	34
Opposites.....	12	12	11	12	12
Word-building..	10	10	9	10	12
Substitution...	68	64	60	58	57

<sup>1</sup> From Hollingworth and Poffenberger's *Applied Psychology*, by permission of D. Appleton and Company, New York.

Chapter IV) there is a drop of about 16 per cent from the earlier to the later ages. It appears that this particular mental characteristic shows the effects of senescence earlier than the characteristics involved in the first three tests. It is quite possible that there are many other tests in which the scores will be appreciably changed by age when still well within the age limits involved in industrial work.

**Sensory defects.** Marked sensory defects will make the tests worthless. If verbal directions are used, a partially deaf person will be at a distinct disadvantage and may not be able to understand at all what is wanted. In a test administered individually or to a small group, this condition will probably be noted by the examiner and proper adjustments made. Visual defects are likewise serious. If a man holds the paper in different positions with apparent effort to focus his eyes upon it, the fact is obvious. Some men likewise will mention the fact that they left their glasses at home. But defects of a less marked degree may nevertheless have an effect in decreasing a person's speed of reading so that his test score does not reflect his actual ability. In lieu of ocular examination it is well to ask the subject if he has ever had any trouble with his eyes.

**Literacy.** One final point should be considered regarding the subjects — their literacy. Many of the tests used are verbal in character and require that the subject be able to read. The general status of the subjects can probably be ascertained from the employment department or by a casual survey of application blanks. If

## THE SUBJECTS USED IN EVALUATING TESTS 183

the literacy of a subject is so low as to handicap him in taking the ordinary tests, recourse must be had to tests of the performance type or at least to tests that involve only isolated numbers or letters.

### SUMMARY

In selecting the subjects on whom to evaluate the tests that are to be used in predicting the potential efficiency of prospective employees, it is possible to use either applicants or present employees. The applicants have the advantage that they are typical of the group on whom the tests are ultimately to be used. They likewise have maximum incentive in taking the tests, as their job is more or less at stake. Their results are not influenced by previous experience on the job in question. Their outstanding disadvantage, however, is that it is necessary to wait weeks or months until they have demonstrated their ability or inability in the job before the criterion is available. If present employees are to be used the criterion is available at once. Incentive is not as effectively insured and special effort must be made to provide it. It is also possible that some of the tests are actually influenced by experience on the job, i.e., measure some acquired proficiency rather than innate capacity. This can be ascertained by correlating test scores with length of service and if a high correlation is found such tests may well be discarded. The most common procedure is to use present employees as subjects on whom to standardize the tests.

In some jobs that do not have many employees it is well to use the entire group as subjects in developing predictive methods for that job. In other cases it is necessary to make a selection or sampling of the group. It is theoretically desirable to have a sufficiently large number so that the addition of others will not appreciably change the results. It is advisable to have the sampling done by the experimenter rather than by the foreman. He may make it purely random or, knowing the criterion in advance, may make it in conformity with a normal frequency curve, i.e., comprising more men of average ability and a decreasing number as the extremes of ability are approached. It is important for the sample to cover the entire range of ability.

Several other factors should be considered in some situations.

Workers of the two sexes should preferably be evaluated separately. Test scores of persons past middle life should be interpreted with caution because some aspects of mental efficiency decline with age. Likewise tests given to persons in their teens should be carefully scrutinized because of the demonstrated fact that proficiency in some tests reaches its maximum at as early an age as 13, while with others maximum proficiency does not occur till 18 or 19. Defects of vision or hearing may vitiate the results if they pass unnoticed. Finally, the literacy of the subjects often imposes limitations on the types of test used.

## CHAPTER VIII

### SPECIAL CAPACITY TESTS: TOTAL MENTAL SITUATION

PREVIOUS chapters have discussed the methods of devising and administering mental tests, of obtaining the criterion and selecting subjects upon whom to standardize the tests. We now turn to the application of the foregoing principles to the actual employment situation. It will be recalled that in the introductory chapter the fundamental principle was laid down that it is necessary to evaluate the tests devised for predicting success in some particular occupation by comparing efficiency in the tests with efficiency in the occupation. This procedure must be followed with every occupation for which tests are desired.

While this general principle applies throughout all such problems, there is a considerable difference between jobs in their mental requirements and the corresponding types of test that prove successful. Recurring to the classification given in Chapter IV, we may subdivide the problem into (1) tests or measures of capacity or aptitude and (2) tests or measures of proficiency. The former functions are presumably innate and the latter acquired. In the former we are concerned with such inborn capacities as attention, memory, or intelligence in so far as they make a man a potentially good tire-builder, even though he has never seen a roller or core, while in the latter we are endeavoring to measure a man's present ability as a carpenter or his ability in some other trade that he has learned. The first of these problems, that of innate capacity or occupational potentiality, is the larger problem. It may be further subdivided on the basis of special capacity, such as attention, memory, reaction time, and general capacity or intelligence.

In devising tests of special mental capacity for predicting vocational aptitude, there are two common methods of approach from the standpoint of the selection of tests. On the one hand, the entire mental situation involved in the job may be reproduced in a single test. For instance, with street-car motormen it is possible to

arrange a continually changing visual environment to which the person must make appropriate reactions and keep about the same attitude of sustained attention and of discrimination between important and unimportant things that is involved in actually driving a car. On the other hand, the work may be analyzed into the mental components involved and these measured separately. For instance, an aviator may need good sense of equilibrium, quick reactions, and emotional stability. It is possible to measure these components and combine them subsequently into a single score. The present chapter will be concerned primarily with the first of these approaches, namely, reproducing the total mental situation involved in the job. It will also discuss certain preliminary steps and certain statistical treatments of results that are equally applicable in approaching the problem by way of the mental components of the job. This latter procedure, however, will be presented in detail in Chapter IX.

#### PRELIMINARY PROCEDURE FOR TEST RESEARCH

**Establishment of *rapport* with those in authority.** The foregoing discussions of tests and criteria have been of rather general and theoretical scope. A few words are in order regarding the preliminary steps that may be of importance to a psychologist embarking upon a practical program of personnel research in an industrial concern. It is unwise to enter the office at eight o'clock some Monday morning, give the stenographer some blanks to mimeograph, and send a rating blank to the foreman of the wood-heeling department requesting him to write in the names of his men and return the ratings by noon. The foreman may not appreciate what it is all about and be unwilling to cooperate, and the psychologist may be making a mistake in starting with the wood-heelers or even in studying them at all. The first introductory step in undertaking such a project is to get in proper *rapport* with those in authority in the concern. While a few of the executives who have been instrumental in authorizing the work may know something about it, the other executives and foremen may be distinctly at a loss to understand what is going on. It is well, then, to meet all those who will be in any way concerned with the project and whose cooperation

may be needed, either individually or in groups, and to discuss with them methods and plans. Perhaps the personal experience of the writer in initiating such a project will afford a sufficiently typical illustration, although, of course, one must adapt himself to the local situation. The executives, foremen, inspectors, and practically all those in authority whose coöperation was needed had an informal club which was designed to promote *esprit de corps* and to provide a forum for discussing various problems of the industry. Fortunately this club had a regular meeting the day after the writer arrived and he was scheduled as the speaker. It was thus possible to present the matter to the entire group at once. A brief discussion was given of psychology in general and certain notions regarding pseudo-psychology — which was the only kind of which the majority had ever heard — were dispelled at the outset. The importance of psychology in industry and particularly in employment was brought out with illustrative material from the army work and from other industrial studies. The experimental standpoint was stressed and the fact that tests are not devised by inspiration or omniscience and immediately put into the employment office, but that they must be tried out on employees whose ability in the job is known. This led up to the importance of the estimates of occupational ability that would be needed. The men were shown how the value of the entire project depended on the accuracy of these estimates and hence that they themselves were just as important as the psychologist. The cards were laid on the table, so to speak, and the men were shown how they could vitiate the entire project if they wished. Furthermore, it was brought out that the final results would not be infallible and that the most the tests could hope to do would be to predict the probable success of a man who scored high in the tests, and that while in the long run more of those with high scores will doubtless be successful than of those with low scores, there are bound to be occasional instances in which a man who gives every indication of the requisite capacity for good work fails to come up to expectations. It is important to drive home this point to the business man because he is prone to attach undue significance to a dramatic instance in which the tests fail, and give insufficient consideration to those instances in which test score and occupational



success coincide. This tendency to note the striking, and neglect the typical instances of relations between things is one of the outstanding fallacies in popular reasoning and accounts for many of our superstitions and other groundless beliefs. The foregoing presentation of principles was, of course, made in terminology with which the men were familiar, and the psychologist made a very definite effort to "sell" them the theory and also to sell them himself and his program.

The next step in establishing *rapport* was to have all these men take some tests. The management authorized and arranged for this and was incidentally interested in the mental status of all those in executive positions. But the testing had the further advantage of familiarizing them all with the nature of tests. Some had been thinking in terms of bumps on the head, and to take a pencil and themselves to mark a test blank was illuminating. A rather wide range of tests (comprising about two hours) was given them in small groups, starting at the top of the organization and working down. Consequently, if any one demurred he could be told that "so-and-so" higher up than he had previously been tested. Thereafter, when any one of them sent in his subordinates for examination, he was in a position to give them a general notion of the situation they were to meet, with a view to allaying their fears, and he could assure them that he himself had been through it and survived.

**Personal orientation.** So much for getting the proper *rapport*. The next introductory step was for the psychologist to get oriented himself. The most obvious necessity was to go over the plant thoroughly and to become familiar with the operations in all parts. An extensive trip was made under the supervision of one familiar with the entire plant and then considerable time was spent individually in going about and observing various departments more in detail. In this way it was possible to get one's bearings and to obtain some notion as to where the big problems lay that were amenable to psychological solution. For instance, in the mill room men were breaking up and washing crude rubber. This was obviously a job that required little intelligence and merely enough ability to keep one's hands out of the rollers. It appeared doubtful whether tests for that kind of work would be profitable. On the other hand,

pressmen were noted watching gauges rather closely and controlling the pressure while the tires were being cured. This appeared to call for rather special ability and might form an interesting psychological problem. However, further observation revealed that there were only a half-dozen such men in the plant — obviously too few for statistical purposes. The same thing was true of the calender men feeding the machines that attached a layer of rubber to a layer of cloth. The building and finishing departments, however, each had some fifty or more workers performing fairly complicated operations. This looked offhand like a place where profitable psychological research might be conducted. Truckmen were hauling trucks about the plant with rubber in various stages of compounding or tires in various stages of being built on heavy cores. This work apparently could be done by an animal under a little supervision and did not appear to offer a profitable field for investigation. A lot of youngsters were handing out stock, i.e., looking at a tag and then going to the place where the stock was "booked" and bringing back the proper kind — apparently a job that required some memory with enough general ability to follow directions. There were likewise a considerable number of these workers and here again there seemed to be promise of interesting experimental results. So it went throughout the plant. This kind of observation was valuable in getting the proper point of view, in becoming familiar with all parts of the plant and with the terminology, and especially in ascertaining where psychological methods might be applied with some hope of success. It also promoted the proper attitude on the part of all concerned as the psychologist was seen about and his face became familiar so that there would be less emotional disturbance later when men came to be tested. Moreover, it was in the line with morale to discuss problems with foremen on the actual spot where the problems arose. A few other things were important in getting this personal orientation. Data on labor turnover were run down to see where the most serious difficulties from the standpoint of the management lay. The methods of keeping the payroll and the like were studied with a view to the possibility of getting production criteria. The general organization was studied to find who was responsible for various

things, so that when anything was needed either in the line of supplies or executive orders the proper person could be approached.

The foregoing preliminary considerations proved valuable in the case in question. Doubtless they are more or less typical of what would exist in other places. The psychologist must, of course, adapt himself to the circumstances, and procedure that will be successful with one group of men and one type of organization may fail with another. But he must definitely strive to get the members of the organization to see his point of view and to understand at what he is driving so that they will cooperate. He must also be familiar with the plant so that he will not make false starts or mistakes due to ignorance of operations or conditions.

#### ANALYZING THE JOB WITH A VIEW TO SELECTING TESTS

**Observation of workers.** Before it is possible to reproduce the total mental situation involved in the job or to devise tests for its mental components, it is obviously essential to find out what elements are involved in the job from the mental and motor standpoint. There are several lines of approach for making such analysis. The psychologist may obtain a good deal of this information by simply observing workmen at the job. If his training has been adequate, he is able to observe persons more closely than is the ordinary individual. In a psychological clinic an examiner must be on the alert for significant involuntary movements and traces of emotional instability. He becomes accustomed to going beyond the mere verbal responses of the patient and to interpreting certain mental aspects in the light of what he does. Laboratory training in experimenting on normal individuals will also help in this respect, for the psychologist must watch his subject as closely as the chemist watches the reactions in the test-tube. Hence he will be in a position to note whether the workers are performing their tasks automatically or with apparent conscious effort, whether they have to attend only to one thing or to distribute their attention to a number of things simultaneously, whether they have to exercise a certain amount of judgment or whether the decisions are made for them, whether they apparently take advantage of any rhythm in the operation and so on. He will also note various other more ob-

jective aspects of the work, such as whether it involves large or small muscle groups, whether the time taken to make the motions is critical, whether the men use near or distant vision and whether they have to remember numbers, symbols, or facts. The psychologist, in short, by virtue of his training will watch the man rather than the machine and will try to analyze the operation from the standpoint of the operator. This type of observation then will contribute materially to the discovery of the factors which it is most essential to include in the mental testing project.

**Questioning workers.** Further information may often be obtained from the workers themselves. It is usually worth while, unless the individuals are of a very low intellectual order, to talk their work over with them. It is certainly out of the question to give them a questionnaire regarding their work, but a personal interview will often yield valuable results. In an interview it is possible to adapt procedure to the circumstances and to follow a lead when it arises. For instance, one may ask a workman what he thinks about during his work. It may then develop that certain aspects of the job require a good deal of attention. A worker may be apparently making very effective reactions to the machine without effort, but will testify that he has to keep "on his toes every minute," thus indicating the importance of ability to sustain attention. A particular delicate motion may be made with apparent ease, but the man will state that he has to "brace himself from his feet up every time" in order to hit the hole, indicating that a high degree of coordination is necessary. If the worker is asked what he finds most difficult about his work, he may, for instance, suggest that he has to be very quick or else "get left." The value of the information obtained in this way will depend largely on the skill of the person interviewing the worker. It calls for tact and patience because workers are liable to be skeptical and hesitate to talk about their work. The interviewer must also be sufficiently familiar with the operation to talk to the workman in his own language. If skillfully handled, however, this procedure will often yield information of value.

**Questioning executives or foremen.** The information obtained from the workers should be supplemented by questioning their superiors. These latter naturally give the objective sort of in-

formation that the psychologist himself gets from observing the workers, but the executives and foremen may have been observing the workers for years and have discovered aspects of the job which would escape the psychologist in his briefer observation. Furthermore, foremen have often worked at the job themselves previously and can see it from both standpoints. It may be feasible to ask the foreman or executive to make out a list of the traits which he considers essential for work of this sort, or to describe in detail the qualities of a good worker. It may sometimes be desirable to go to the foreman with a list of traits or qualities and ask him to indicate the ones that are most important in the present connection. Such a list should not be taken, of course, at its face value, but may very well be used as a starting-point for further personal discussion. It is sometimes illuminating to get two foremen together and have one "hire" the other, i.e., stage an employment interview. This may yield information of positive value or quite the reverse, as in an experience of the author's in which it developed that a machinist's proficiency is judged largely by the kind of set of tools that he claims to possess. A good starting-point for interviewing a foreman is to ask what is most frequently the trouble with a man who fails to make good at the particular type of work in question. The reply that he does not put his mind on his work suggests the desirability of using some sort of a test for sustained attention. The statement that he is too slow indicates the possibility of using a test that involves reaction time. The procedure to be followed depends upon the individual foreman. As a rule there will be less difficulty in interviewing him than in interviewing the workers, for he will be on the inside and will understand the nature of the whole project.

**Personal experience.** It is often advisable for the psychologist to try the job himself. He may have had some laboratory training in self-observation and thus be able to note subtle aspects of the mental state during the job that would otherwise be overlooked. He will see for himself just how difficult it is to coordinate, what initial adjustments of attention are necessary, what judgments are involved, how far estimates of space or time are crucial, and to what extent quickness of reaction is essential. It will be illuminating, anyway, to see the job from the inside. The psychologist who did

an extensive piece of research in connection with methods of selecting taxicab drivers started in by driving a cab himself for a few weeks. Although it may not be desirable to pursue a job till a high degree of skill is obtained, the undertaking of the initial stages of learning may contribute information of value.

**Previous job analysis.** This preliminary psychological analysis is not exactly the same as the procedure of job analysis which is discussed in Chapter XV. The latter is often more comprehensive than that above outlined and involves such things as the following: The exact duties involved, the working hours, the general conditions of work with reference to posture, temperature, and hazards, physical qualifications such as strength, vision, or hearing, education, previous experience, amount of judgment and supervision involved, as well as such things as speed or accuracy. The ordinary job analysis is made by trained interviewers who go over specific topics with workers or foremen and write up the occupational description on the basis of these interviews. Such specifications often involve, however, many things that are at least by implication of a psychological character, and if such specifications are available for an occupation the investigator beginning a project of developing mental tests for that occupation will doubtless find the specifications of considerable value. They may not be entirely adequate, especially if made by a person without psychological training, but they will probably call attention to some facts that the psychologist might overlook in his own preliminary analysis. He will naturally go more directly at the psychological aspects of the job. Perhaps the ideal situation is that in which both kinds of information are used. If the job analysis results are available, the psychologist may well take them as a starting-point and then proceed still further with the specifically psychological aspects of the job. When his final conclusion is reached as to the mental factors involved in the occupational situation, he is ready to develop a test or tests for those factors.

By way of illustrating the analysis of a specific job and developing a test therefor, a study of hand-fed dial machine operators will be cited. (337, 112.) Other tests of this type will be described later in the chapter. The present one will be discussed more in detail in

order to illustrate the various steps involved in the developmental procedure. These hand-feed dial machines have a series of holes in a rotating table. These holes must be kept filled with material that is to be stamped. The operator has a supply of material and, as the empty holes pass by the point nearest to him, he inserts the material in the proper holes. Analysis of this operation by the procedures mentioned above indicated that it seemed to involve rather sustained attention toward a particular point on the dial and the adjacent portions in the direction from which the empty holes came. It also involved rather close coördination of eye and hand in hitting the hole accurately. Moreover, there was a sort of bodily rhythm in feeding the machine as it was driven automatically and the holes passed at constant rate. In this case it proved feasible to devise a single test which reproduced this whole mental situation. In other instances, to be described in the next chapter, it is preferable to measure the different mental components of the job separately.

#### REPRODUCTION OF THE TOTAL MENTAL SITUATION

**Simplicity vs. complexity.** After analysis such as the foregoing the next step is to reproduce the whole situation as far as possible. This necessitates some device that will get the subject into about the same mental attitude that he would have in the actual job. In the development of apparatus for such purposes there are a number of things to bear in mind. One should not complicate the apparatus needlessly. While this does no harm it is unnecessary and involves useless expense and effort. Frequently a rather simple device can be made which will give exactly the same effect, as would a much more complicated machine, as far as the mental state of the subject is concerned. This point is particularly pertinent because the apparatus at the outset is purely experimental and may be scrapped if it fails to give results which correlate with the criterion.

**Adaptation of existing apparatus.** Another somewhat similar point is that it is often possible to use or adapt existing apparatus rather than to develop something entirely new. The psychologist has opportunity to exercise considerable ingenuity in adapting such things as an old phonograph or typewriter to his purposes.

Recurring to the illustration of hand-feed dial machine operators, a large metal disc was mounted on the chassis of a phonograph which drove it at constant speed. Near the margin of this disc were two slots of regulable size. Beneath one point where these slots passed was a funnel. The subject was provided with steel balls which he dropped through the slot into the funnel where they were recorded by a mechanical counter. It was possible to vary the speed of revolution or the size of the opening. Balls which did not go through the slot rolled to another opening where they were recorded separately. It will be seen that the whole device was relatively simple and utilized some existing piece of apparatus rather than building up an entirely new mechanism for supporting the rotating slot. This was highly desirable, particularly at the outset, because there was no assurance at all that the apparatus would be permanently used.

**Subjective vs. objective similarity.** It should not be assumed from the foregoing that in reproducing the total mental situation it is necessary actually to reproduce the job on a miniature scale. It is the subjective rather than the objective similarity between test and job that is important. For instance, in a test for street-car motormen it was not necessary to provide a toy car and toy pedestrians. Red and black numbers appearing in different positions at a window in the apparatus were used to produce the rapidly changing mental situation involved in driving a car, and the mental aspect was quite similar to that involved in actual practice, although objectively the materials were dissimilar.

**Fool-proof.** In devising such a test, the principles discussed in Chapter V should of course be observed. A test for total mental situation generally involves some form of apparatus rather than a printed blank and is perforce an individual test. It is essential with this sort of test to insure its fool-proof character technically. There should be no way in which the subject can beat the apparatus either through cleverness or stupidity. If, for instance, a motion is to be made in only one direction, this can be insured by using a ratchet arrangement so the other direction is actually impossible. In the test for dial-machine operators the size of the slot was such that two balls could not be inserted at once. Any that failed to go



through into the funnel rolled to one side and were caught by an apron. Such points as these must be carefully observed in order to avoid a subject's making a higher score than he ought or a totally unreliable score by some means which circumvents the experimental situation.

**Objective scoring.** Again, in this type of test the method of scoring should receive especial consideration. It is undesirable to have the performance one in which the examiner must judge qualitatively by observing the subject, but rather one which yields some quantitative measure of proficiency. This will usually be in the form of quantity of work done per unit time or time taken for a certain amount of work. In the above example a mechanical counter in the neck of the funnel recorded the number of balls successfully dropped into it; another counter on the disc recorded the number of revolutions or the maximum number of balls that could have been dropped through it. It was then possible to note the actual per cent of efficiency in directly quantitative form. This was far more satisfactory than it would have been to eliminate the counters and determine by watching the subject whether his performance was good or otherwise.

#### GIVING THE TESTS TO WORKERS WHOSE CRITERION IS AVAILABLE

**Separate laboratory vs. laboratory in the shop.** When tests for the total mental situation have been devised as the result of analysis the next step is to evaluate them by giving them to a group of subjects and correlating test scores with the criterion. One of the first problems that arises here as well as in the test for mental components (*infra*) is to decide under what conditions the tests should be given. There are two tendencies in this respect — to have a laboratory or testing-room separate from the factory proper or to do the testing under conditions closely approaching those in the shop — perhaps in the shop itself. Psychologists differ in their predilection for these two methods. The advantage of the former is that in a separate laboratory there is more quiet and less distraction by the noise of machinery. It is possible to set up the apparatus in more permanent form. The room may be locked so that things will be undisturbed. The room may be likewise fitted with tables for test-

ing small groups. In favor of the latter alternative it is urged that the workers are in a more natural situation and less apt to have emotional disturbances such as they would get in coming to a separate laboratory. Moreover, they can be obtained somewhat more easily, as it is only a few steps from their work and they can be taken when they come to a good stopping-place. Sometimes a small room adjoining the shop is equipped for this purpose or sometimes a small portable affair of beaver board is constructed which can be set up in any department and is large enough for a table, chairs, and requisite testing equipment. (337, 62.) This factor of avoiding undue nervousness on the part of the subject is perhaps more critical with women operatives than with men, particularly if the examiner is a man. Some go so far as to have the laboratory in the shop and to leave the door open so that the forelady can "chaperon."

The writer is inclined to favor the separate laboratory rather than the shop laboratory. The advantages of quiet, more flexible experimental set-up and permanency outweigh for him the more natural surroundings for the subject and the convenience of the subjects in going to and fro. The writer has tested many women operatives in an isolated laboratory and never appreciated the need of a chaperon. If the proper introductory steps have been taken so that the entire plant knows something about what is going on, there is little shock when a person comes to be tested if proper tact is used. The usual procedure, moreover, is to precede the tests proper with a "shock absorber" test which is not scored at all, but which merely gets the subject into the proper attitude and takes off the novelty of the situation. Furthermore, inasmuch as the tests are ultimately to be used to examine applicants in the relatively quiet employment office rather than in the factory, if present employees are tested in relatively quiet conditions their results and those of subsequent applicants should be comparable.

Mention should be made again at this point of the desirability of determining a test's reliability (cf. p. 136). If this has not already been established, it is desirable in the present situation either to give the test twice (in a different form if necessary) to each subject or else to provide for separate evaluation of different parts of the test.

It is then possible to note whether those who make a high score in one part do likewise in the other, thus ascertaining whether the test is reliable.

#### CORRELATION OF TEST SCORE WITH CRITERION

After the tests have been given to a group of subjects and scored, the next step is to compare the test scores with the criterion in order to determine whether those who are efficient in the tests are efficient in the job and *vice versa*. This makes it possible to state whether the tests are valid and can be subsequently used with applicants to predict their occupational efficiency. The correlation procedure to be discussed in connection with tests for total situation will be equally applicable to the tests for components to be discussed later.

Various methods are available for indicating this correspondence between test scores and criterion. The method that is to be used will be determined somewhat by the form in which the criterion is obtained. If it is possible merely to have the workers grouped into two classes, good and poor, or two classes at the extremes of ability, about all that can be done is to compute the average test score made by each group. If the good workers make appreciably higher scores on the average than do the poor workers, this indicates something. But the result is not in a form that will enable one to make a very definite prediction of occupational efficiency. If an applicant is given the test, about the only statement that can be made is that his test score is a certain amount above or below the average made by the good workers and it will be impossible to state how big a chance is being taken in hiring him. What is ultimately wanted is some indication of the *probability* of occupational success on the basis of the test scores. This goal necessitates the computation of correlation coefficients. The present procedure is so far inferior to correlation procedure that it will not be discussed further. Every effort should be made to obtain the criterion in such a form that correlations can be computed.

**Rank-difference method.** The technique of correlation has already been mentioned (p. 29). It aims to derive a quantitative expression of the tendency for two variables such as test and job to

be related so that those who score highly in one are apt to score highly in the other and *vice versa*. One common method of correlation consists of ranking the individuals with respect to each variable and then noting the differences in rank. We may call the best person in the test 1, the next best person 2, the third best 3, and similarly the best one in the job may be called 1, the next best 2, the third best 3. If then for each individual we note the difference between the two ranks assigned him we may get some notion of the correlation of the two variables. For instance, if all the differences are small it shows that a person is ranked about the same in both test and job, while if the differences are large this indicates considerable discrepancy between persons' rankings in test and job. From these differences it is then possible by appropriate formulæ to compute a coefficient which will indicate the closeness of relation between efficiency in the test and efficiency in the job. Several examples are worked out by this method in Appendix I (examples I-V), which illustrate not only the method of computation, but also how the correlation coefficient expresses quantitatively the closeness of the relation. When there is a perfect relation the coefficient is 1, while if there is no relation at all it is 0. It may even take on negative values as large as  $-1$ , indicating that the better a person is in one respect the worse he is in the other.

**Products-moments method.** The rank-difference method has the drawback mentioned in the previous discussion of ranks that it assumes that the first person is just as superior to the second as the second is to the third. This often obscures extreme tendencies that ought to be considered. The correlation procedure devised to meet this contingency is called the "products-moments" (i.e., products of deviations) method. It determines essentially whether deviations from the average in one variable are accompanied by corresponding deviations in the other, i.e., whether a person is about as far above the average in one respect as he is in the other and *vice versa*. It is necessary to compute these deviations from the average for each individual measure, to get the products of each pair of deviations and to perform some other computations. An example is worked out by this "products-moments" method in Appendix I, example VI.

**Scatter plot.** The foregoing methods are often quite tedious when a considerable number of individuals are involved. Moreover, it is sometimes desirable, with a set of test scores which are being correlated with a criterion, to observe whether there are many individuals who are poor in the job, but who nevertheless do well in the tests. From a small correlation coefficient alone it may not be evident whether the lowering of the coefficient is due to such individuals or to persons who are good in the job, but poor in the test, or to both kinds. Hence it is sometimes desirable to present the relation graphically. This makes it possible to discover at a glance any particularly anomalous cases, such as the predominance of persons who are poor in the test, but good in the job. It is possible also with the data in this graphic form to compute a "products-moments" coefficient by short-cut methods.

		Criterion									
		1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Test Score	46-50									C	
	41-45								R		G
	36-40								H J		
	31-35						I		L		
	26-30			F		B		P			
	21-25						K				
	16-20			E	A Q		M				
	11-15				O						
	6-10	D	T	S							
	1-5		N								

FIG. 4. SCATTER PLOT FOR CORRELATING TEST AND CRITERION

This graphic method involves the construction of a "scatter plot." The procedure is illustrated in Table XVI and Figure 4. The table gives the test scores and criterion scores for 20 workmen A, B, C, etc. It may be noted that the test scores range approximately from 1 to 50. In this particular instance this range is

TABLE XVI. DATA TO ILLUSTRATE THE SCATTER-PLOT METHOD OF CORRELATION

WORKMAN	TEST SCORE	CRITERION
A	17	36
B	28	44
C	49	88
D	6	4
E	20	22
F	29	24
G	42	97
H	36	73
I	32	54
J	40	79
K	22	54
L	32	77
M	16	52
N	3	12
O	14	33
P	28	65
Q	18	37
R	41	72
S	9	25
T	8	13

divided into 10 classes and the rows of the chart laid off accordingly. For instance, in the bottom row are to be placed men who score between 1 and 5 in the test; in the next to the bottom row are to be placed men who score between 6 and 10 in the test. Similarly, the criterion scores range from approximately 1 to 100 and are likewise

divided into 10 classes and the columns of the chart labeled accordingly. In the first column are to be located men who score between 1 and 10 in the criterion; in the next column men who score between 11 and 20. The choice of exactly 10 classes is not essential. In actual practice anything from 10 to 20 classes proves satisfactory. We are now ready to plot the data of Table XVI in the appropriate rows and columns of Figure 4. Workman A has a test score of 17, which locates him in the row marked 16-20; his criterion is 36, which places him in the column headed 31-40; there is only one compartment of the table determined by this row and this column, and consequently A is written in this compartment. Similarly, B's test score of 28 puts him in the 26-30 row; his criterion of 44 puts him in the 41-50 column; only one compartment is determined by this row and column, and B is written in that compartment. In the same way all the other individuals are plotted. In actual practice letters or names are not entered in the chart, but merely an "x" or check mark of some sort. If many entries occur in a given compartment, they are subsequently replaced by a single figure which gives the total number of entries.

A glance at Figure 4 shows a rather definite tendency for the entries to scatter more or less along a diagonal line — from the lower left to the upper right corner. Those in the lower left corner are poor in both test and criterion, while those in the upper right are good in both respects. In general, the farther to the right a person is located, the nearer the top of the chart he is located; i.e., the better he is in the criterion, the better he is in test score. This indicates a high correlation between the two variables. With a scatter plot like this, it is possible by short-cut methods to compute the actual products-moments correlation coefficient. In the present instance that coefficient is .90.

By way of comparison two other scatter plots are given in Figure 5. The class intervals are not indicated, but merely the general trend of the distribution shown. Each dot represents an individual. The chart at the left involves a negative correlation. It is to be noted that the entries scatter roughly along a diagonal line from the upper left to the lower right corner. This means that those who are high in test score tend to be low in the criterion and

*vice versa*. The general trend is exactly the reverse of that shown in the preceding figure and yields a large negative correlation, whereas the former case yielded a large positive correlation. The other chart in Figure 5 shows the kind of a scatter plot resulting from data with a very small correlation. It is obvious that the entries are scattered at random in the plot and there is no tendency at all for high scores in one variable to go with high scores in the other.

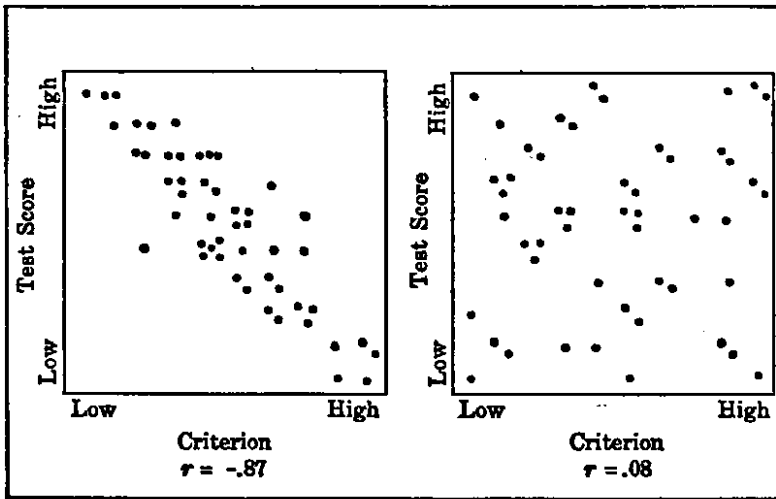


FIG. 5. SCATTER PLOTS FOR A LARGE NEGATIVE AND FOR A SMALL CORRELATION

Methods such as the foregoing afford the best approach to the problem of the relation of test scores to criterion. The particular method used will vary with the nature of the data and the statistical or computing equipment available. Work of the highest order usually demands a products-moments correlation coefficient in the final evaluation of the two variables. Whether this is computed by dealing with the actual scores or by plotting them first does not make so much difference unless one is interested in locating anomalous cases. The main point is to obtain the best possible quantitative expression of the validity of the test, i.e., the tendency to which high test scores go along with good ability in the job and



*vice versa*. In the example of dial-machine operators above mentioned the correlation between scores in the test and piece-work earnings was approximately .50. This is only a fair correlation, but indicates some validity for the test.

**Regression equation.** It is rather conventional practice after a correlation has been worked out to derive a "regression equation." This is simply an equation that expresses criterion in terms of the test and is of the form:

$$X = bY + C$$

where  $X$  is the criterion or ability in the job and  $Y$  is test score and  $C$  is a constant term. The  $b$  term is proportional<sup>1</sup> to the correlation between criterion and test. This equation gives us the best estimate that we can get as to ability in the job on the basis of the test. If, for instance, the equation proved to be  $X = .6 Y + 20$  and a given applicant for a job scored 80 points in the test, we merely substitute 80 for  $Y$  in the equation, thus:  $X = .6 \times 80 + 20$  or  $X = 68$ . This means that 68 points in the criterion is the best prediction we can make as to his ability. The equation comes out, of course, in whatever terms have been used to obtain the criterion. If the latter was obtained in earnings per hour, the equation will predict the most probable earnings per hour, while if it was in terms or ratings on a linear scale, the prediction will be in those terms. The prediction cannot be made, of course, with absolute certainty, but the equation gives us the best prediction that can be made with the available data. The closeness of prediction that can be made with correlations of different magnitude will be discussed later in the chapter.

#### LIMITATIONS OF TOTAL SITUATION METHOD

A test for total mental situation has only one serious limitation as compared with tests for the mental components to be discussed

<sup>1</sup> The detailed formula is:

$$X - M_x = r \frac{\sigma_x}{\sigma_y} (Y - M_y)$$

where  $\sigma_x$  is the standard deviation of the criterion scores and  $\sigma_y$  the standard deviation of test scores,  $M_x$  is the mean or average of the criterion scores and  $M_y$  the average of the test scores.

subsequently, and that is the fact that if the test shows a low correlation with the criterion all the work has been wasted. If a rather complicated device has been constructed and workmen have been obtained to take the test and then the final score fails to correlate with ability in the job, it is necessary to begin all over again. This is often rather difficult to do in a particular industrial organization. If the men are called in again for further tests, they naturally wonder why they could not have had these new tests at the outset. It gives the psychologist the appearance of not knowing what he is about. In the method of components, on the other hand, a considerable number of tests are given and separately correlated with the criterion. Those that show a low correlation can be scrapped. Some, however, will probably show an appreciable correlation and can be used without the necessity of calling the men back for further examination. The former procedure amounts to keeping all the eggs in one basket, while the latter distributes them so that the prospect of an utter catastrophe is less. There are situations, however, in which it is almost certain in advance that a test can be devised which will reproduce the situation and show an appreciable correlation with the criterion. In other cases it may be possible to obtain the subjects repeatedly without inconvenience. Sometimes the test for total mental situation may be given along with some tests for mental components. In all such cases the method is justified.

#### EXAMPLES

**Motormen.** A few other examples of tests designed to reproduce the total mental situation involved in the job will be described. Some of these, however, do not embody all of the principles laid down above. The classical example is Münsterberg's test for motormen. (408, 69). Analysis of the process of driving a street-car indicated that the operator had to be alert to the changing visual situation before him and had to discriminate between the lengthwise traffic which was relatively harmless and the crosswise traffic which was a potential source of danger. Moreover, he had to note the character of this traffic — pedestrian, horse, or automobile — in order to determine the probable rate of motion relative

to the track. This total situation was reproduced by a relatively simple device. An endless belt (a strip of velvet with cards attached to it) moved across a narrow window from the top toward the bottom of the window. Along the middle of this belt were two lines representing the track. There were four spaces half an inch wide, indicated by lighter lines, on each side of the track. In these spaces numbers 1, 2, and 3 were distributed. Black numbers indicated lengthwise traffic and red numbers indicated crosswise traffic. Number 1 represented a pedestrian and, if red, was supposed to move one space between the time it appeared and the time the car reached the point where the pedestrian was crossing. Thus, a number 1 in either zone adjacent to the track was dangerous, while in any of the other zones it could be neglected because the car would pass before the pedestrian reached the track. Similarly, number 2 represented a horse-drawn vehicle and was supposed to move two spaces between the time of appearance and the time the car arrived. Consequently, such a number in the second zone from the track would be dangerous, but in other zones would be harmless, as the vehicle would either get across or not reach the track at all in the allotted time. In like manner number 3 represented an automobile and was supposed to go three spaces in the crucial time. Consequently, these numbers in the third zone from the track would be significant, but in other zones could be neglected, as they would cross in time. The subject taking the test moved the belt by turning a crank himself and called out at every dangerous thing which appeared at the window. He had thus to discriminate the colors (direction of traffic), the numbers (nature of traffic), and the position of the numbers (location of traffic). In this way he was in much the same attitude of attention and discrimination he would be in actually driving a car, although he was operating a small device and reacting to mere abstract symbols. The time taken was recorded with a stop-watch and the number of mistakes noted. These were combined into an arbitrary score and compared with the criterion — number of accidents in the man's service record. The method of scoring was not ideal and the statistical treatment was not rigorous. Münsterberg, however, reports a "far-reaching correspondence between efficiency in the experiment and efficiency in the actual service."

**Gun-pointers.** Another example of a test for total mental situation is that devised for gun-pointers. (142.) The mental situation involved is fairly obvious. The man has to keep his gun trained on the target by appropriate manipulation, and at the proper time fire or give a firing signal. It involves an eye-hand coordination with a moving target. To reproduce this situation a small target was moved in a horizontal plane by means of a piston falling in oil. A lever a meter long had a small telescope mounted on it. The subject operated the lever to keep the hairline in the telescope trained on the target. A pencil attached to the lever traced on a rotating drum the actual motion of the lever. It was possible subsequently to clamp the target to the lever and trace on the record the actual course of the target. These two tracings were then compared by inspection. It would not be possible without much more elaborate technique to obtain a quantitative record. However, using mere qualitative impression there was sufficient correspondence between test results and efficiency in actual gun-pointing to warrant installation of the methods on several battleships.

**Aërial observers.** A test for aerial observers was developed along these same lines. (141.) The observer is particularly concerned with the location of different objects on the terrain below him. To measure this capacity a set of aluminum plates was provided, each containing circular holes arranged in an irregular pattern. These holes were covered with colored paper and illuminated from behind. The subject controlled this illumination himself by operating a key. He turned on the light, studied the pattern until he thought he knew it, and then immediately tried to reproduce it. If his effort was incorrect, he was instructed to study it further. Record was made of the time taken for study and for reproduction as well as the number of trials before a correct response was made. Then this pattern was shown several times during a series of others and the discriminative reaction time measured, i.e., the time in fractions of a second required to decide whether the pattern was the one previously seen or was a different one. The observers were classed into three groups on the basis of their actual efficiency in aerial observation and different aspects of the test compared with this criterion.

The clearest results were obtained with the discriminative reaction time. The average time taken by the best observers to discriminate that the pattern was the same as the one learned at the outset was .75 seconds, while for the other aerial observers it was 1.09 seconds. Similarly, the times taken to discriminate that the pattern was different were .92 and .98 respectively for the two groups. The scores of the two groups overlapped considerably, but there was nevertheless some indication of the validity of the test.

**Miscellaneous examples.** Several attempts have been made to reproduce in the laboratory the general situation involved in flying an airplane. For instance, a chair was mounted so that it could rotate within a rectangular frame. This frame rotated in another larger one and this, in turn, in a third. By levers which controlled motors driving the shafts of the various frames, the subject who was strapped in the chair could turn it in any desired direction. Other devices involved complicated levers for tipping the chair or reacting to various signals or targets which were presented (cf. 529). No results are available in which such tests were compared with actual efficiency in flying an airplane.

A test for prospective lathe operators involved two large screws similar to the feed screws on a lathe mounted at right angles. (443.) The free end of these screws carried a member with a writing point. The subject, by turning these screws simultaneously, attempted to make the writing point follow a prescribed pathway and deviations from this pathway could be noted. This test shows fair agreement with proficiency in a college course in shop practice and when combined with some other tests yields a correlation of .55.

Among the various tests for telephone operators have been some that involved a miniature switchboard. (349.) A more elaborate test for telegraphers involved learning five letters of the Morse code. These, then, were to be recognized in various combinations, at various rates and along with distracting sounds. (343.)

A number of complicated elaborations of Münsterberg's earlier test for motormen have been made. The signals to which the subject reacted were more varied as to location and distance. The subject was required to react in several different ways, sometimes with actual controls similar to those used on a car and

sometimes with levers that involved about the same muscle groups without actually duplicating the standard controls. Other variations consisted of additional emergency signals which required a particular reaction or of a rattle which could be introduced into the sound of a motor and which the subject had to detect quickly. The entire results were recorded automatically so that they could be analyzed at leisure with especial reference to errors. When performance in the test is compared with managers' judgments in one form or another the results in various studies have been shown graphically to be of some significance or have been reported as "satisfactory." (218, 504, 513, 635, 661.)

#### CRITICAL SCORES

After tests for total mental situation or for mental components of the job have been devised and given to operatives of known ability and the final correlation of test or weighted sum of tests with criterion determined, the problem arises as to just how the tests are to be used for employment purposes.

It would not do to tell the employment manager, or the man who did the hiring: "Here is the regression equation to use for your general clerical help:  $X = 3Y + 20$ ." The reaction of the average employment man to such a statement could easily be predicted. His real problem is whether or not to hire an applicant, and if mental test scores are available for that man they must be interpreted in easily understandable fashion. Although we may know that the test gives a fairly valid prediction of probable success in the job, we wish to know the probable success of the particular man. This involves the notion of critical scores, i.e., the determination of a score below which a person should not be hired because of lack of promise of success. This score must be based on the probability that persons who fall above it will succeed or that those who fall below it will fail. It is thus important to consider the tests from the standpoint of probability, for, as suggested previously, tests seldom predict with absolute certainty. Consequently, there are bound to be certain instances in which a person seems promising on the basis of the test but fails to come up to expectations. These dramatic instances are apt to catch the attention of the management.

to the exclusion of the cases of correspondence. Hence it is desirable to emphasize this probability aspect of the critical score as soon as it is determined.

There are two general methods of determining these critical scores. One is to compute the theoretical score in the job from the regression equation (cf. p. 204) and perhaps to generalize this by dividing the range of test scores into a number of classes and computing the chances of occupational success for each class. The other is to compute the per cent of occupational successes and failures above and below certain test scores. This procedure may often be handled easily by graphic methods.

**Probability of occupational success predicted from a regression equation.** As stated above, if we give a prospective employee a vocational test and substitute his score in the regression equation, the result gives us the best possible prediction of his ability on the job that can be made with that test. The same reasoning applies if we give him a series of tests and work out a more complicated regression equation embodying scores in several tests (cf. next chapter). The prediction is, as previously mentioned, in the terms in which the criterion was originally obtained such as salary or ratings by foremen on a millimeter scale. But while the value given by the regression equation is the most probable salary or standing in the foremen's estimation, we are not sure that it is absolutely correct.

When dealing with probabilities we do not hit the mark every time. To draw an analogy from another field, if four coins are tossed the most probable result is two heads and two tails, but if they are tossed repeatedly this result will not always be obtained. Sometimes there will be one head or one tail and occasionally all heads or all tails. In fact, if the coins are tossed 1600 times there will be approximately 100 cases of all heads, 100 cases of all tails, 400 cases of one head and three tails, 400 cases of one tail and three heads, and 600 cases of two heads and two tails. Thus the best guess as to what will occur in any given toss is two heads, but one cannot be absolutely sure of tossing it. However, one would rather bet that the result of any given toss would be two heads than to bet that it would be three or four heads. That is, the actual values that would

be obtained if the event were repeated many times would average around the most probable value without always coinciding with it. This same principle applies to the probable value of the criterion computed from a regression equation involving mental test scores. If we had a large number of men who made exactly the same score in the tests and the regression equation indicated a value of \$60 as the most probable salary, if we put these men to work and after they had learned the job tabulated their actual salaries they would average about \$60, but some would be a little more and some a little less. While perhaps the majority would receive \$60, there would be some who would receive \$65 and about the same number who would receive \$55. There would probably be others, fewer in number, receiving \$70 and \$50, with fewer still receiving lower or higher wages than these. In other words, the actual salaries if plotted in the form of a distribution curve would give the normal type of frequency distribution (cf. Figure 1, p. 163) with the high point at \$60, the most probable value, and with salaries above and below that occurring symmetrically and with decreasing frequency the more they deviate from \$60 in either direction. That is to say, there is a certain error involved in estimating one variable from other correlated variables — in this case in estimating success in the job from the tests. This is termed the "standard error of estimate" and is computed by the formula  $\sigma\sqrt{1-r^2}$  where  $\sigma$  is the standard deviation of the thing we are trying to predict, i.e., ability in the job, and  $r$  is the correlation between job and test. It will be seen that the larger the correlation, the smaller the error of estimate. With a high correlation between criterion and test one can hit the mark rather closely in predicting vocational ability on the basis of the test. With a low correlation there is a big chance of falling rather wide of the mark.

It is possible to use these facts when a given worker is tested in order to determine his probable success. We can take his job score computed from the regression equation as the most probable value and then make a distribution curve with this as the average and with the standard error of estimate operating as the standard deviation of the curve. This is exactly the same procedure as that outlined in Chapter VI. We can plot a normal frequency curve if we



know the average and the standard deviation of the measures or estimates. Then, if we lay off the base line of this curve in units of the standard deviation, we can determine what proportion of the cases fall between any assigned limits. (Cf. Figure 2, p. 165, and the accompanying discussion.) The "cases" in this instance are estimates of job score on the basis of the regression equation. Consider the smaller curve in Figure 2. Suppose that \$60 is the most probable salary computed from the regression equation and that the standard error of estimate is \$20. If this same result was obtained for a great many men, it is probable that with about 34 per cent the salary would actually (after the man had learned the job) prove to be between \$60 and \$80, because between the average and a value greater or less than it by an amount equal to the standard deviation fall 34 per cent of the cases. Similarly, in 48 per cent of the cases the actual salary would fall between \$60 and \$100 and likewise we should expect 34 per cent between \$40 and \$60 and 48 per cent between \$20 and \$60. Putting it in another way, if a man's most probable salary is found to be \$60, the chances are 34 out of 100 that his actual salary will be between \$60 and \$80 and they are 48 out of 100 that his actual salary will be between \$60 and \$100. On the other hand, there are 34 chances out of 100 that he will earn between \$40 and \$60 and 48 chances that he will make between \$20 and \$60.

We can then decide whether to "take a chance" in hiring the man. Suppose any one who will ultimately earn less than \$40 is undesirable, the chances in the present case are 16 out of 100 (i.e., 50-34) that the man will be in that class. Now, suppose in another set of tests which have a higher correlation with the criterion, the most probable salary is likewise \$60, but the standard error of estimate is only \$10. The probability is then 48 out of 100 that the man will actually earn between \$40 and \$60, because \$40 is less than the average and by amount equal to twice the standard deviation and there are only 2 chances out of 100 of his being in the undesirable class of those who make less than \$40. Thus, with this higher correlation and smaller standard error of estimate, we are taking only 2 chances out of 100 of getting a poor man, while with the lower correlation and higher standard error of estimate we are taking 16

chances out of 100. This shows the desirability of high correlations on which to base the prediction of probable success in the job.

Tables for probable distribution of occupational ability on the basis of test scores. The foregoing method of taking the test scores for each individual applicant and then determining the probable distribution of his success in the job is often too cumbersome for the ordinary employment procedure. The same principles may be used for a given job in working out a general table which shows for various ranges of test scores the chances of attaining different degrees of proficiency in the job. Suppose we have 1000 workmen and we divide them into 10 groups on the basis of the tests — the best 100, the next best 100, etc., down to the worst 100. Now, suppose we likewise divide them on the basis of their ability in the job into the best 100, the next best 100, etc. It is then possible to take the best 100 in the tests and note how many of them are in the best tenth in the job, how many in the next best tenth, etc., down to how many are in the worst tenth. Then we can take the second 100 in the tests and see how many of them are in the highest tenth in the job, how many in the next highest tenth in the job, etc. While it is possible, if given enough cases, to construct such a table empirically from the actual data, it is likewise possible, knowing the correlation between test and criterion, to work out such a table in general that will hold for predicting any variable on the basis of another provided they have the correlation indicated. This latter procedure is perhaps somewhat better because such a table, worked out, for instance, for a correlation of .60, can be used in any subsequent vocational situation in which test and criterion correlate to the extent of .60.

A few such typical distributions are given in Table XVII. They show the probability of occupational success as predicted from test scores when the correlations are .00, .50, .60, .70, .80, and 1.00. The rows in the table, indicated by roman numerals, give the 10 different degrees of ability manifested in test scores, while the columns, indicated by capital letters, give the 10 degrees of occupational ability. For instance, consider the correlation of .70. Suppose the 1000 men are divided into 10 classes on the basis of their test scores. Class I represents the best 100 and class II the next best 100.

TABLE XVII. FOR INTERPRETING CORRELATION COEFFICIENTS OF DIFFERENT MAGNITUDE

I, II, III, etc., indicate successive deciles (tenths) of test scores; A, B, C, etc., indicate successive deciles of vocational ability.

$r = .00$

	A	B	C	D	E	F	G	H	J	K
I.....	10	10	10	10	10	10	10	10	10	10
II.....	10	10	10	10	10	10	10	10	10	10
III.....	10	10	10	10	10	10	10	10	10	10
IV.....	10	10	10	10	10	10	10	10	10	10
V.....	10	10	10	10	10	10	10	10	10	10
VI.....	10	10	10	10	10	10	10	10	10	10
VII.....	10	10	10	10	10	10	10	10	10	10
VIII.....	10	10	10	10	10	10	10	10	10	10
IX.....	10	10	10	10	10	10	10	10	10	10
X.....	10	10	10	10	10	10	10	10	10	10

$r = .50$

	A	B	C	D	E	F	G	H	J	K
I.....	32	19	14	10	8	6	5	3	2	1
II.....	19	16	14	12	11	9	7	6	4	2
III.....	14	14	13	12	11	10	9	8	6	3
IV.....	10	12	12	12	12	11	10	9	7	5
V.....	8	11	11	12	11	11	11	10	9	6
VI.....	6	9	10	11	11	11	12	11	11	8
VII.....	5	7	9	10	11	12	12	12	12	10
VIII.....	3	6	8	9	10	11	12	13	14	14
IX.....	2	4	6	7	9	11	12	14	16	19
X.....	1	2	3	5	6	8	10	14	19	32

SPECIAL CAPACITY TESTS

TABLE XVII. FOR INTERPRETING CORRELATION COEFFICIENTS OF DIFFERENT MAGNITUDE (continued)

I, II, III, etc., indicate successive deciles (tenths) of test scores; A, B, C, etc., indicate successive deciles of vocational ability.

$r = .60$

	A	B	C	D	E	F	G	H	J	K
I.....	39	20	14	10	7	4	3	2	1	0
II.....	20	19	16	13	10	8	6	4	3	1
III.....	14	16	15	13	12	10	8	6	4	2
IV.....	10	13	13	13	12	12	10	8	6	3
V.....	7	10	12	12	13	12	12	10	8	4
VI.....	4	8	10	12	12	13	12	12	10	7
VII.....	3	6	8	10	12	12	13	13	13	10
VIII.....	2	4	6	8	10	12	13	15	16	14
IX.....	1	3	4	6	8	10	13	16	19	20
X.....	0	1	2	3	4	7	10	14	20	39

$r = .70$

	A	B	C	D	E	F	G	H	J	K
I.....	47	22	13	8	5	3	1	1	0	0
II.....	22	22	17	13	10	7	5	3	1	0
III.....	13	17	17	15	12	10	7	5	3	1
IV.....	8	13	15	15	14	12	10	7	5	1
V.....	5	10	12	14	14	13	12	10	7	3
VI.....	3	7	10	12	13	14	14	12	10	5
VII.....	1	5	7	10	12	14	15	15	13	8
VIII.....	1	3	5	7	10	12	15	17	17	13
IX.....	0	1	3	5	7	10	13	17	22	22
X.....	0	0	1	1	3	5	8	13	22	47



Similarly, class A represents the best 100 in the job and class B the next best 100 in the job. The table shows that of those in class I in the test there will probably be 47 in class A in the job, 22 in class B in the job, 13 in class C, while there will be none in classes J or K. By contrast, with a correlation of .50, of the men in class I we find only 32 in class A, 19 in class B, and actually find several in classes J and K. Obviously, with a higher correlation there is less chance that those with high test scores will do poorly in the job. The extreme cases of correlations of .00 and 1.00 show this in a still more marked fashion.

**Prediction of success of individual applicant.** Instead of interpreting the table in terms of the number of the group who will have different degrees of ability in the job, we may equally well use it for a given man who falls in any particular tenth in the tests to predict the chances of his falling in any of the 10 classes in the job. This inference from the proportion of a group to the chances of an individual is a common one. If an actuary finds that 30 people of your age and status out of 100 die before they are 60 years of age, the chances are 30 out of 100 that you will die within that time. Similarly, if test and criterion correlate to the extent of .70; any man whose test score falls among the highest 10 per cent of test scores stands 47 chances out of 100 of being in the highest 10 per cent in the job, 22 chances out of 100 of being in the next highest 10 per cent in the job, etc.

Knowing, then, the correlation between a particular set of tests and the criterion, it is possible by this procedure to work out a distribution like those in Table XVII. Then, when an applicant is tested it is possible to note in which class of test score he falls and compute his probability of attaining the various degrees of occupational success. The determining of a critical score then involves merely the consideration of how big a chance the management wishes to take.

This may be illustrated by recurring to our example of 1000 men distributed as in Table XVII. Suppose that the workmen on the job at the present time who are in the lowest 10 per cent — i.e., class K — on the basis of occupational proficiency are manifestly unsatisfactory and it is desired in future to hire as few as possible of

this sort. Suppose the correlation between test and criterion is .70. Referring to the distribution for this correlation, if we hire from 1000 applicants only the 100 best men in the test scores — i.e., if we establish a critical score between classes I and II — we shall obviously get no one from class K. The same will be true of those in class II in the tests, so if the critical score is established between classes II and III no one in K will be hired. If, however, the line is drawn between classes III and IV — i.e., if the 300 best men in the tests are hired — there will be one of them in the unsatisfactory vocational class K. If the line is drawn between IV and V, there will be 2 out of the 400 men unsatisfactory or .5 per cent, and if it is drawn between VI and VII, there will be 10 out of the 600 unsatisfactory or 1.6 per cent. Or suppose that both classes J and K — i.e., the lowest 20 per cent in occupational ability — are to be avoided with a correlation of .70 again. If the critical score is established between classes II and III — i.e., if the best 200 men are hired — only one of them will be undesirable, i.e., .5 per cent; if the line is drawn between IV and V, there will be 11 such out of the 400, i.e., 2.7 per cent; while if it is drawn between VI and VII, there will be 36 out of the 600 unsatisfactory, i.e., 6 per cent. In this way it is possible to see just what per cent of those hired who fall above a certain critical score in the test will be unsatisfactory in the job.

**Justification of efforts to raise the correlation between test and criterion.** If we now carry through this same reasoning with coefficients of different magnitude, we can see how the value of the tests in eliminating unsatisfactory workers depends on the size of the correlation between test and criterion. Take, for instance, the above problem of eliminating all individuals in classes J and K — the lowest 20 per cent in occupational ability whom we will call “unsatisfactory” workers. Suppose the labor market is such that we are enabled to hire the best 20 per cent in the tests, i.e., we place the critical score between II and III. If the correlation is .50, we shall by this procedure hire 4.5 per cent of our workers who are unsatisfactory; if the correlation is .60, we shall get only 2.5 per cent such workers, i.e., only about half as many, and if the correlation is .70, we shall be accepting only .5 per cent, while if it is .80, we shall get

none at all. These figures appear in the second column of Table XVIII. In other words, if in this particular instance we can de-

TABLE XVIII. PER CENT OF UNSATISFACTORY WORKERS (CLASSES J AND K) THAT WILL BE SELECTED IF CRITICAL TEST SCORE IS DRAWN BELOW THE CLASS INDICATED

<i>r</i>	II	IV	VI
.50	4.5	7.5	10.7
.60	2.5	5.0	8.2
.70	.5	2.7	6.0
.80	.0	.7	3.5

vised a test which correlates .60 with the criterion rather than .50, we almost double our ability to eliminate these unsatisfactory workers, while if we can find one with a correlation of .70 we make only one ninth as many mistakes as with a correlation of .50.

This type of example makes clear the justification of the effort to obtain a test with as high a correlation as possible. Similar implications will be brought out in the next chapter, where a number of tests are used and considerable labor involved in "weighting" them statistically with a view to increasing the correlation between the sum of the tests and the criterion. The saving will not always be of exactly the magnitude indicated in the present example, since it will depend on where the critical score is drawn and the proportion that it is desired to eliminate. In the above example, if the critical score is drawn between classes IV and V or between VI and VII, the results are somewhat different. These facts are embodied in the remaining columns of Table XVIII. The figures in the columns marked IV and VI are obtained in exactly the same manner as those in column II described above. In all these cases the higher correlation very manifestly eliminates more of the undesirable workers.

Simplification of the practical problem of prediction. After a distribution such as those in Table XVII has been worked out, it may be desirable to simplify the administration somewhat. Instead of stating the classes in the tests merely as "highest tenth,"



etc., it is perhaps preferable to state their limits in terms of actual scores. Furthermore, the best three tenths in the criterion may be arbitrarily designated "good workers," the next three tenths "average," and the next three tenths "poor," while the lowest may be termed "very poor." On this basis with a correlation of .70 this will mean that of those in the best tenth in the test 82 per cent will be "good" in the job (47 + 22 + 13), 16 per cent average, 2 per cent poor, and none very poor; of those in the second best tenth in the test 61 per cent will be good, 30 per cent average, 9 per cent poor, and none very poor. To cite a practical statement of this sort Table XIX was used for interpreting the psychological tests used in the Air Service. The weighted sums of test scores were in this case divided into 8 instead of 10 classes, but the general theory was exactly like that discussed above. The table is in a form that is probably more understandable to the layman than the general decile tables like those in Table XVII.

**TABLE XIX. SIGNIFICANCE OF FINAL SCORE IN AIR SERVICE TESTS OF APTITUDE FOR FLYING**

By a "good flyer" is meant the top three tenths of cadets as they are now.  
 By an "average flyer" is meant the next three tenths.  
 By a "poor flyer" is meant the next three tenths.  
 By a "very poor flyer" is meant the bottom tenth of cadets as they are now.

Of 100 cadets scoring:

75 or better.....	85	will be good,	14	average,	1	poor,	0	very poor
50 to 74.....	70	" " "	24	" 6	" 0	" "	" "	
25 to 49.....	52	" " "	35	" 12	" 1	" "	" "	
0 to 24.....	32	" " "	41	" 25	" 2	" "	" "	
0 to -24.....	16	" " "	38	" 40	" 6	" "	" "	
-25 to -49.....	7	" " "	27	" 50	" 16	" "	" "	
-50 to -74.....	2	" " "	16	" 50	" 32	" "	" "	
-75 or lower.....	1	" " "	8	" 40	" 51	" "	" "	

**Graphic methods.** In lieu of this consideration of probable success computed theoretically from the correlation coefficient, simpler graphic methods are often used. If the criterion consists of only a few groups of occupational ability, such as good, average, and poor, it is possible to plot the test scores of individuals in the three groups and see where the line can be drawn with the least possible overlapping of the groups. This procedure is illustrated in Figure 6. It shows the determination of a critical score for pre-

dicting success in agricultural engineering. (105.) The weighted test scores are laid off along the base line and each individual represented by a square above the appropriate score. The individuals who were considered good by their instructors are plotted in the top section of the chart; those who were rated average in the middle

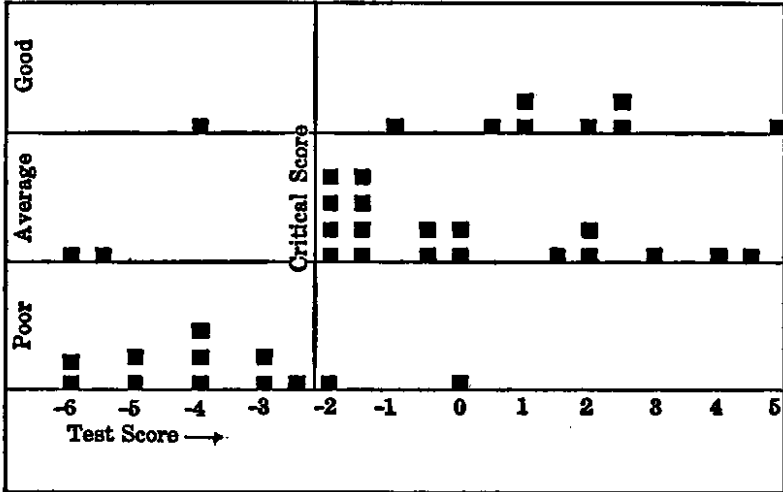


FIG. 6. GRAPHIC DETERMINATION OF CRITICAL SCORE

section, and those who were poor in the lowest section. After the persons are plotted in this fashion, it is necessary to determine by inspection where to draw a line that will make the sharpest division between those in the poor section and those in the other sections. In the present instance, if the line is drawn between  $-2.5$  and  $-2$  this makes a fairly good division. There are only 2 of the poor engineers who do better than this critical score, so that there is not a very large chance of admitting inferior individuals if such a score is used for vocational advice. On the other hand, there are only 3 of the average or good engineers who fall below this score, so if it were used there would be very few desirable individuals ruled out along with the undesirables.

When the criterion is available in more detailed form and the graphic method is to be used, a scatter plot similar to Figure 4

(p. 200) may be constructed. Recurring to that figure, suppose that practical considerations indicate that workers with a criterion score (salary, foremen's estimates, or what not) of fewer than 41 points are undesirable. If a vertical line is drawn between the 31-40 and the 41-50 class, we wish to hire as few persons as possible who are to the left of this line, but to employ as many as possible to the right of this line. The problem is to draw a horizontal line such that most of those below it will be to the left of the first line and *vice versa*. If we draw such a line between the 16-20 and the 21-25 classes of test score — i.e., if we employ no one who scores less than 21 — we shall obviously be eliminating most of the undesirable men. There is only one such (F) who falls above this critical score. On the other hand, there is only one of the desirable men (M) who will be eliminated by this procedure. Hence a critical score of 21 points in the test may well be adopted. Persons scoring less than this have little chance of coming up to the requirements of the occupation, while most of those scoring above this amount will qualify.

The determination of critical scores, as above described, depends somewhat on the relation between the number of applicants for work of a given sort and the number of vacancies. If the situation is such that there are no more applicants than vacancies so that little selection can be made, it is a question of ruling out only the very worst prospects and hence a rather low critical score must be used. On the other hand, when the number of applicants far exceeds the number of vacancies so that only a small per cent can be hired, it is to the benefit of all concerned to have those hired with the best promise of success. In this case a rather high critical score may be set.

#### SUMMARY

In embarking upon a program of personnel research in an industrial concern, at least two preliminary steps are desirable. The psychologist must, in the first place, establish *rapport* with those in authority so that they will be ready to coöperate in every way necessary. To this end the general nature of the project may well be explained to them and they should be shown their own importance therein. It is also well to familiarize them with test proce-

dure by having them take some tests themselves. In the second place, the psychologist must orient himself in the organization. He will need to be familiar with the different operations and with the terminology. He may likewise locate departments where there appears to be the greatest need for research and where conditions are favorable for obtaining valid results.

In devising tests of special mental capacity for predicting vocational aptitude, there are two common methods of approach — reproducing the total mental situation involved in the job or analyzing the operation into its mental components and testing these components separately. In either instance it is necessary to make a preliminary analysis of the mental aspects of the job. It is also necessary to give the test or tests to workers and to correlate the score or scores with the criterion. To analyze the mental aspects of the job, it may be well to observe workers carefully, actually to try the job and observe one's own experiences, to discuss the requisites with foremen and executives with especial reference to the distinguishing features of efficient and inefficient workers, or to use as a starting-point a job analysis that has previously been systematically conducted.

In devising the test for total mental situation, it is wise to avoid undue complexity, because the apparatus is at the outset purely experimental and may later be scrapped. The test need not necessarily be a miniature of the job, because it is the subjective rather than the objective similarity that is important. It should, however, be technically fool-proof and yield an objective score.

The next step is to give the test to subjects whose ability in the job is known. The testing may be done in a separate laboratory or in a screened portion of the factory. The former affords more quiet and allows more flexible and permanent equipment, while the latter is more natural and convenient for the subjects. The emotional factor, however, can usually be controlled by giving a "shock absorber" test preceding the crucial series.

After the test has been given to a group of workers, it is necessary to correlate the scores with the criterion. This may be done by appropriate formulæ which consider the differences between each subject's rank in the test and rank in the criterion, or which involve

the product of each man's deviation from the average test score and his deviation from the average criterion score, or the data may be plotted with test scores on one axis and criterion scores on the other. In any instance the magnitude of the correlation coefficient indicates the validity of the test. It is also possible to work out a regression equation which expresses criterion in terms of test score and gives the best prediction that can be made of the man's ability on the job with that particular test.

The test for total mental situation has one serious limitation. If its correlation with the criterion proves to be small, the work has been practically wasted and it is necessary to start again. It is often difficult or embarrassing to have the same subjects return later for further examination.

Various examples of such tests were cited. The situation for hand-feed dial-machine operators was reproduced by a rotating disc containing a hole through which steel balls were dropped by the subject. A test for motormen involved an endless belt with a track in the middle passing an opening in the apparatus. Numbers at various positions relative to the track had various significance and the subject reacted accordingly. Gun-pointers looked through an eye-piece and by a hand lever kept it trained on a moving target. Aerial observers were required to memorize certain patterns of illuminated points that were flashed on electrically. The validity of these and other tests that were cited was sufficiently high to warrant their practical use.

After the tests for total mental situation or for mental components have been devised and correlated with the criterion, it is necessary to determine a critical score. This is a score such that persons falling below it will receive unfavorable consideration for employment. The essential thing from the employment standpoint is the probability that the applicant will be a successful worker after adequate training. This may be best determined on the basis of the regression equation. This equation expresses ability in the job in terms of test scores, and by substituting a given applicant's scores in the equation it is possible to determine his probable ability in the job. This prediction, however, is not absolute and his actual ability may deviate somewhat from the pre-

dicted. But the higher the correlation of test and criterion, the closer will the actual ability fall to that predicted from the regression equation. It is also possible to compute the chances of the actual ability falling within any particular limits above or below the predicted.

To simplify the administration, it is possible to work out for any given correlation a general table showing for various ranges of test scores the chances of attaining various degrees of occupational proficiency. We may arbitrarily call certain ranges of ability good, average, and poor, and then state the probability of a given applicant being a good, average, or poor workman. The employment department can then decide where to draw the line for a given set of tests on the basis of how large a chance it wishes to take in hiring applicants. This line is the critical score.

Instead of using a regression equation and computing the probability of occupational success, graphical methods may be used for rougher determination. If the workers are grouped into a few classes of occupational ability, distribution curves may be plotted for each group and lines drawn to make the best separation between the groups with the least overlapping. If the results are put in the form of a scatter plot with test on one axis and criterion on the other, a line may be drawn to indicate the limits of occupational ability below which it is undesirable to employ a man and then another line at right angles may be drawn by inspection so as to have most of those below a certain test score in the inferior occupational group.

## CHAPTER IX

### SPECIAL CAPACITY TESTS: THE MENTAL COMPONENTS OF THE JOB

As suggested at the outset of the preceding chapter, there are two leading methods of approach to the problem of tests of special mental capacity for predicting vocational aptitude. The first of these — reproducing the total mental situation — was the main topic of Chapter VIII. The other may be termed the method of mental components. The essential feature of the method is the determination of what mental factors are involved in the job and then the devising of tests which measure these separately as far as possible. Instead of one test with one final score for the whole mental situation involved in the job, we have a number of tests for the different factors involved and combine them into a single score. Moreover, it is possible to determine the best method of combining them in the particular situation in order to get the most valid prediction.

#### PRELIMINARY SELECTION OF TESTS

**Analysis of job into its mental components.** In order to devise tests for the mental components of the occupation, it is necessary, of course, to have some notion of what those components are. This analytic procedure has already been described in some detail in the preceding chapter. The psychologist may find it profitable to observe the men at work, to talk with them, perhaps to try the work himself and to discuss with foremen or executives the characteristics of the good and poor workers at this job. If a job analysis has been made, this will often afford valuable insight and give the psychologist a starting-point for his own analysis. This procedure yields a number of mental factors that are presumably involved in the case of a person working at this occupation.

**Devising of tests measuring these components.** The next step is to select or devise mental tests which roughly measure these factors. As stated earlier, there is probably no test that measures

a single factor and discrete mental factors may not exist anyway. However, there are occupations in which good attention is obviously a requisite and occupations which patently require memory and there are tests which to a very considerable extent give a measure of ability to concentrate and ability to remember quite apart from the fact that the test may measure other additional things. If such tests are selected in the light of preliminary analysis, there is a much greater chance of obtaining some high correlations with the criterion than if tests are selected at random. The number of tests included in this preliminary selection depends largely on the length of time for which each subject will be available. The more tests used, the greater the probability of finding some which are valid, just as the more shots fired at a target, the greater the chance of hitting the bull's-eye. If the analysis indicates relatively few factors that seem obvious, it is well to employ several tests that roughly measure each of these factors, such as several attention tests, several memory tests, or several motor coördination tests, because, while two attention tests may be to quite an extent similar, they may nevertheless vary sufficiently to catch some particular mental aspect that is significant in the job in question.

As an illustration of this procedure we may consider the job of finishing automobile tires. (102.) The tire comes to the finisher with several plies of fabric already built on an iron core. He puts it on a frame so that he can spin it by hand in a horizontal plane, places plies of gum stock on the tread and rolls them down with hand rollers. In some cases a line is traced around the tire with a pair of dividers and stock has to be applied with its edge along this line. The workmen testified that they had to "keep their mind on the work" in order to be successful. The foreman said that the men who fell down on the job were "too slow." Careful observation of the men at work suggested that they required a rather distributed attention, needed to be able to sustain their attention, i.e., concentrate for a considerable time without a break, and that quick reaction time, good motor coördination, and ability to judge distances were essential. It was feasible to have each employee for one hour's examination. Consequently, tests that roughly measured the above factors were selected for one hour's work.



The material and method conformed to the principles laid down in Chapter V.

Fifteen tests were selected for trial. Test 1 consisted of tracing through a series of rectangular patterns between two lines about one eighth of an inch apart, keeping time with a metronome. Test 2 involved tapping with a metal ring on the tip of the forefinger and making contact alternately with brass plates mounted one above the other two inches apart. Test 3 was similar to example 2 in Chapter IV (*supra*). Test 4 was a modification of the one described as example 1 in Chapter IV. Test 5 involved aiming at a target with a pencil — the target being at arm's length and the hand brought back to the shoulder between attempts to hit a series of crosses on the target, keeping time with a metronome. Test 6 employed a series of shotgun shells made up in different weights to determine the smallest difference in weight that the subject could discriminate. In test 7 the subject traced a line with a pencil, then drew a line of the same length while the copy and his hand were covered with a screen. Test 8 involved cancelling pairs of numbers as in example 11, Chapter IV. Test 9 comprised a page of disconnected, unspaced letters — the subject to underline groups of adjacent letters that formed a word. Test 10 involved finding consecutive numbers that were arranged at random (example 13, Chapter IV). Test 11 was a substitution test similar to example 14 (*supra*). Test 12 comprised a series of mazes like example 15 (*supra*). Test 13 was simple visual reaction time, i.e., the fraction of a second taken on the average to release a simple electric contact when a stimulus object moved. (Cf. example 19, Chapter IV.) Test 14 involved watching a moving target which passed in front of an opening in a screen and then continued at the same rate while invisible. The subject was required to stop the invisible target at some designated point by pressing a key and the discrepancy between actual position and designated position noted. Test 15 comprised a series of lines each accompanied by a short line. The subject determined without measuring how many times the shorter was contained in the longer.

The tests were not actually given in the above order. Those requiring considerable mental effort were interspersed with those

which were more motor in character in order to obviate any undue fatigue. Each test moreover was divided into two installments. The subject went through the first installments of each test and then went through the second installments in the same order. This made it possible to compute the reliability of the tests.

#### FINAL SELECTION OF TESTS

After the tests have been selected, they are given to a group of workers whose occupational ability is known in the same fashion described above in connection with the test for mental components. The subsequent procedure, however, is somewhat different. In the former case the single test yielded a single score and it was simply a question of the extent to which this score correlated with the criterion. In the present case there are many tests and many scores. It is a question of selecting the best ones from this set of tests and discarding the others. Moreover, some of the tests that are retained correlate more highly with the criterion than do others and consequently should play a larger part in determining the final score. If one test is twice as good as another, it should play twice as important a rôle in the final prediction. If a third test is five times as good as the first, due allowance should be made. In such a case, in order to get the best prediction of a man's ability in the job on the basis of the tests, it would be necessary to multiply his score in the first test by 1, his score in the second test by 2, and his score in the third test by 5. This procedure of determining some constant number by which to multiply each score is termed "weighting" the tests. It can be shown that if a set of test scores are weighted properly, they will give a better prediction of occupational ability than if they are combined in some other fashion. It generally develops that a relatively small number of tests properly weighted will give as good a prediction as a large number. Furthermore, the statistical labor in weighting more than ten tests in the best possible manner is very considerable indeed. All these facts, then, indicate the desirability of selecting from the large group of original tests a smaller number that are retained for more intensive study.

Preliminary correlation of each test with the criterion. This se-

lection of the most promising tests is usually made by some preliminary sort of correlation procedure. This will vary somewhat with the circumstances and with the form of the available data. It is not always necessary in this preliminary sorting to employ the

TABLE XX. CORRELATION OF PRELIMINARY TESTS WITH ABILITY IN FINISHING TIRES

Test 1.....	.21
Test 2.....	-.09 .02
*Test 3.....	.31
Test 4.....	-.02 -.05
Test 5.....	.02
Test 6.....	.01
Test 7.....	-.03 -.07
*Test 8.....	.49
*Test 9.....	.30
*Test 10.....	.52
*Test 11.....	.35
Test 12.....	.07
*Test 13.....	-.38
Test 14.....	.10
Test 15.....	.06

relatively laborious products-moments correlation coefficient, for the purpose is merely to eliminate the tests that are absolutely hopeless. In some instances a comparison of the average score made by a group of the best workers with that made by a group of the worst men will give the desired preliminary information. If the number of workers involved in the study is not too large, the method of rank differences is not especially laborious. With more individuals

in the group it is common practice to make scatter plots as above described and determine by inspection which are the worst tests. If there is no semblance of conformity of the check marks to a diagonal distribution, it is useless to consider that particular test further. From the original list of tests by means of some such methods the worst ones are eliminated and a smaller number, usually not over ten, of the most promising retained for further study.

To continue the previous example of developing tests for tire finishers, about fifty employees were examined. Estimates of foremen and production figures yielded a criterion score for each workman. The correlations of test scores with the criterion were computed by the method of rank differences. The coefficients are given in Table XX. In instances where two correlations are indicated for a given test, there were two methods by which the test was scored that were evaluated separately. Obviously, some of the tests are worthless. Consequently those nine tests with low correlations were scrapped and the other six (indicated by stars in the table) retained for further study.

#### WEIGHTING THE FINAL GROUP OF TESTS

The next step is to determine the proper weight to assign to each of the tests that is retained, i.e., to determine a number by which to multiply scores in that test before totaling into a single combined score. In rough work where correlation coefficients are not available, but where the average test scores made by a group of good workers are compared with the average scores made by a group of poor workers, the difference between these average scores gives some indication of the value of the test. If with one test the good workers do thirty per cent better than the poor, while with another test the good workers excel the poor by sixty per cent, we may roughly say that the second test is twice as good as the first. The only logical weighting then would be to give the former a weight of 1 and the latter a weight of 2. This procedure would doubtless be more satisfactory than weighting the two tests equally.

If the data have been worked out in the form of correlation coefficients, it might seem logical to weight the tests directly in pro-

portion to these coefficients. If one test correlates with the criterion to the extent of .30 and another to the extent of .60, the weights might be 1 and 2. This procedure, however, if several of these tests are to be used, overlooks one very important point, viz., the tests overlap one another in varying degrees. Suppose that memory and attention are actually of equal importance in the job and that two tests of memory and one test of attention are retained and suppose they all correlate equally with the criterion. If they are all added together with equal weight, we are giving twice as much consideration to memory as to attention in the final score and are selecting employees preponderatingly on the basis of memory, whereas attention should receive equal consideration. This procedure obviously is unsound and takes no account of the fact that the two memory tests overlap.

**Correlation of tests with each other.** This overlapping of the tests can readily be determined by correlating the tests with each other. In the above instance, if scores in the first memory test are correlated with corresponding scores in the second, a high correlation coefficient will doubtless be obtained, while the attention test will probably not correlate as highly with either of the memory tests. This indicates from another angle that the attention test should receive greater weight than either of the others because it is measuring more of a unique factor while the others overlap. In cases where the close relation between several tests is not patent from the nature of the tests themselves, this procedure of intercorrelating the tests is very essential. It is then possible to make proper allowance. Knowing these intercorrelations, the next problem is to determine, not merely where allowance must be made in the weighting, but *how much* allowance must be made for the overlapping. A statistician who has had considerable experience in such matters may often by inspection make a rather shrewd guess as to the proper weighting. A technique, however, is available for determining these weights in the best possible manner. This technique is termed "partial correlation." Allusion has already been made to this method in connection with the weighting of speed and accuracy (p. 135). For full discussion of the technique the reader is referred to advanced works on statistics. (272, 692.)

In the present connection effort will be made to present only the general principles and a rudimentary notion of the technique.

**Partial correlation.** The scientist is often interested in determining the relation between two things. The chemist studies this relation between the pressure and volume of a gas, the physicist the relation between current and resistance in a circuit, and the employment psychologist the relation between memory test scores and occupational proficiency. The logical experimental approach to the problem is to change one of the factors under consideration and note whether the other changes and how. The chemist varies the volume of the gas to note what happens to the pressure, the physicist alters the resistance in the circuit and measures the corresponding changes in current, and the psychologist selects workmen of varying proficiency in the job and studies their scores in the memory test. However, the scientist must take account of the presence of other factors which may influence the results. He wants to know the actual or intrinsic relation between the factors under consideration quite apart from other things. If the chemist pays no attention to the temperature of the gas, his findings as to change in pressure are as liable to be due to temperature as to volume. If the physicist fails to consider voltage, he does not know whether the change in current is actually due to resistance. If the psychologist takes no account of other factors, such as attention, it is impossible to tell whether the relation between his test and the criterion is due to memory or to something else. The ideal procedure in such cases is to keep the extraneous factors constant. It is possible for the chemist to keep the temperature mechanically constant throughout his experiment of noting the relation between pressure and volume. The physicist can impress on his circuit a constant voltage while he changes the resistance and measures the current. But there are many problems — and employment psychology faces one of them — where it is impossible *objectively* to keep the extraneous factors constant. It would be difficult, for instance, to find a group of workmen all of whom have the same powers of attention. In such cases it is possible, however, to control these factors *analytically*. Instead of keeping attention constant by selecting a group of workers with identical capacity, it is

possible to test the group that is available and then to determine statistically what the relation between the memory test and the criterion would have been if it had been possible to obtain such a select group with constant attention. This involves the derivation of partial correlation coefficients which indicate, not the *observed* relation between two variables, but the intrinsic relation between them *with other variables kept constant*.

The ordinary correlation coefficient such as we have been studying is often quite misleading because of the presence of other factors besides the two that are correlated. This may well be illustrated by a study made of the relation between hay crop, precipitation, and accumulated temperature. (249, 38.) The figures varied when different parts of the year were considered, but the following set illustrates the principles under discussion. The correlation between crop and precipitation (written  $r_{cp}$ )<sup>1</sup> was .44, an appreciable correlation, i.e., the more it rained the better the crop grew. The correlation between crop and temperature ( $r_{ct}$ ), however, proved to be only .05. This did not look right, for common sense says that things grow better in warm weather. Further computation revealed the fact that the correlation between temperature and precipitation ( $r_{tp}$ ) was  $-.44$ , i.e., as it became warmer it likewise grew drier. This serves to explain the previous coefficient of .05. Some relation actually existed between crop and temperature, but this did not appear in the observed data, because, when the weather became warmer and would naturally tend to increase growth, it likewise became drier and this tendency worked against the other. From the above data it was possible to compute a coefficient of partial correlation (the method will be briefly described below) between crop and temperature with precipitation constant. It was obviously impossible to keep precipitation *physically* constant throughout the years when the observations were made. It was possible, however, to control it *analytically* and to determine what the relation between crop and temperature would have been if the precipitation had been kept constant. This correlation ( $r_{c-t-p}$ )

<sup>1</sup> The common notation in correlation procedure is to write  $r$  (the correlation coefficient) with two subscripts indicating the variables correlated — in this case  $c$  and  $p$ .

proved to be .30.<sup>1</sup> In other words, there was actually some intrinsic relation between crop and temperature, but it was entirely obscured in the objective data because of the presence of the other factor. When it became warmer, things tended to grow (as indicated by the correlation of .30 between crop and temperature with precipitation constant), but it likewise became drier (as indicated by the correlation of  $-.44$  between temperature and precipitation). The net result of these opposed tendencies was no apparent relation between crop and temperature (as indicated by the correlation of .05). This shows how misleading the ordinary type of correlation coefficients sometimes are and how much more illuminating are the partial correlation coefficients.

In scientific study, then, of the relation between two variables, it is desirable to determine their intrinsic relation with other factors as far as possible constant. This principle is particularly pertinent in developing a group of tests for the mental components of the occupation. It is desirable to weight each test, not in accordance with its ordinary correlation with the criterion, but according to its intrinsic relation as revealed by partial correlation. Suppose for instance, that three tests are used and the problem is to find the intrinsic relation between the criterion and the third of these with the others constant. If it were possible to give the first test to 10,000 subjects, we could find all of those who scored equally in this test. Suppose there were 1000 of these individuals. We could then give this 1000 the second test and find perhaps 100 of them who had the same ability in the second test. With this selected 100 who had identical ability in both of the other tests, we could then compute the correlation between the criterion and test 3. We should then have the correlation between the criterion and test 3 with the other factors constant. It is obviously impossible to adopt such procedure in the employment situation; but it is statistically possible to obtain almost the same result if all three tests are given to the limited group of 100.

The technique of computing partial correlation coefficients is

<sup>1</sup> The customary notation with partial correlation is to indicate by the first two subscripts the variables correlated and by the other subscript or subscripts after the period, the variable or variables kept constant.



complicated and laborious. A comparatively brief example is worked out in Appendix III. It is necessary to determine not only the correlation of each test with the criterion, but also the correlation of each test with every other test. These latter correlations are necessary in order to allow for the overlapping of the tests. From these original correlations it is possible to compute partial correlations like  $r_{12.3}$  which indicates the correlation between the criterion (1) and test 2 with test 3 kept constant. From this sort of coefficient, with one test kept constant, it is possible to compute those with two kept constant, like  $r_{12.34}$  which indicates the correlation between the criterion and test 2 with both tests 3 and 4 constant. From these coefficients it is then possible to compute those like  $r_{12.345}$  in which three tests are constant, and so on according to the number of tests involved.

These computations are all made by formulæ like the following:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}$$

Where  $r_{12.3}$  represents the partial correlation between the criterion and test 2 with test 3 constant,  $r_{12}$  is the ordinary correlation between the criterion and test 2,  $r_{13}$  is the correlation between the criterion and test 3, and  $r_{23}$  is the correlation between tests 2 and 3. Suppose  $r_{12} = .70$ ,  $r_{13} = .60$ , and  $r_{23} = .80$ . If we substitute in the formula we have:

$$\begin{aligned} r_{12.3} &= \frac{.70 - .60 \times .80}{\sqrt{1 - (.60)^2} \sqrt{1 - (.80)^2}} = \frac{.70 - .48}{\sqrt{1 - .36} \sqrt{1 - .64}} \\ &= \frac{.22}{\sqrt{.64} \sqrt{.36}} = \frac{.22}{.8 \times .6} = \frac{.22}{.48} = .46 \end{aligned}$$

In the practical situation we are interested in obtaining the largest possible partial coefficients of test and criterion because they enable us to make a better prediction of occupational ability on the basis of the test. Let us consider what things are conducive to large partial correlations. Suppose that in the above example  $r_{12}$  had been .90 instead of .70. The solution of the formula then gives:

$$\begin{aligned}
 r_{12.3} &= \frac{.90 - .60 \times .80}{\sqrt{1 - (.60)^2} \sqrt{1 - (.80)^2}} = \frac{.90 - .48}{\sqrt{1 - .36} \sqrt{1 - .64}} \\
 &= \frac{.42}{\sqrt{.64} \sqrt{.36}} = \frac{.42}{.8 \times .6} = \frac{.42}{.48} = .87
 \end{aligned}$$

The resulting partial correlation of .87 is obviously larger than the .46 obtained in the previous case. This illustrates a fundamental principle, viz., that the larger the ordinary correlation of a test with a criterion, the larger will be its partial correlation with the criterion.

Recurring now to the original example, suppose that  $r_{12}$  and  $r_{13}$  had been the same, but that  $r_{23}$  had been .30 instead of .80. The solution of the formula then gives:

$$\begin{aligned}
 r_{12.3} &= \frac{.70 - .60 \times .30}{\sqrt{1 - (.60)^2} \sqrt{1 - (.30)^2}} = \frac{.70 - .18}{\sqrt{1 - .36} \sqrt{1 - .09}} \\
 &= \frac{.52}{\sqrt{.64} \sqrt{.91}} = \frac{.52}{.8 \times .95} = \frac{.52}{.76} = .68
 \end{aligned}$$

The resulting partial coefficient of .68 is much larger than the .46 obtained previously and it is due entirely to the fact that  $r_{23}$  is smaller. This gives us a second principle, viz., that the smaller the correlation of a given test with another test, the larger will be its partial correlation with the criterion.

These two principles indicate what is necessary if we are to have tests with a high predictive value. If we wish tests which have a high partial correlation with the criterion, those tests are the best whose correlation with the criterion is high and *with the other tests, low*. If two tests show equal correlation with the criterion, but the correlation of the first with the other tests is low, while that of the second with the other tests is high, the former is measuring a more independent factor and its partial correlation coefficient will be higher. It should receive more weight in the final prediction. This, then, is the solution of the problem raised earlier as to how to weight the tests properly in order to obviate the effect of overlapping factors and the danger of giving undue weight to some one factor. The tests are to be weighted, not in accordance with their

ordinary correlation coefficients with the criterion, but in proportion to their partial coefficients with all the other tests held constant. In this way each test gets a weight according to its intrinsic relation with the criterion and it can be shown statistically that this weighting is more valid than any other that may be devised.

**Regression equation.** The actual process of weighting involves the derivation of a regression equation. This is the same sort of thing described in the preceding chapter where ability in the job is expressed in terms of score in the test. In the present case, however, it is expressed in terms of several tests. The equation is of the general form:

$$X_1 = b_{12}X_2 + b_{13}X_3 + b_{14}X_4 + \dots + C$$

in which  $X_1$  represents the criterion,  $X_2$  represents the score in test 2,  $X_3$  represents the score in test 3, etc.,  $b_{12}$  is the weighting for test 2,  $b_{13}$  is the weighting for test 3, etc., and  $C$  is a constant term. The  $b$  terms are, roughly speaking,<sup>1</sup> proportional to the partial correlations —  $b_{12}$  is proportional to the partial correlation of test 2 with the criterion when all of the other tests are constant;  $b_{13}$  is proportional to the partial correlation of test 3 with the criterion when all of the other tests are constant.

TABLE XXI. ILLUSTRATING WEIGHTING TEST SCORES ACCORDING TO A REGRESSION EQUATION

$$X_1 = 7X_2 + 9X_3 + 14$$

	SCORE IN TEST 2	SCORE IN TEST 3	$7X_2$	$9X_3$	CONSTANT TERM	WEIGHTED SUM
Adams.....	12	11	84	99	14	197
Andrews.....	10	8	70	72	14	156
Briggs.....	14	13	98	117	14	229

An illustration of weighting tests according to a regression equation is given in Table XXI. The equation proves to be  $X_1 = 7X_2$

<sup>1</sup> These  $b$  terms also take into account the variability of the different tests. The  $C$  term results from the fact that the equation is first derived in terms of deviations of scores from average score and then transformed into terms of actual test scores.

+  $9X_2$  + 14; that is,  $b_{12}$  is 7,  $b_{13}$  is 9, and  $C$  is 14. The men make the test scores indicated in the first two columns of the table. Adams, for instance, scores 12 in test 2 and 11 in test 3; Andrews scores 10 in test 2, and 8 in test 3. These constitute the  $X_2$  and  $X_3$  values for each man. The weight for  $X_2$  is 7, so each score in the  $X_2$  column has to be multiplied by 7 (cf. the column headed  $7X_2$ ). Similarly, each score in the  $X_3$  column is multiplied by 9 (cf. the column headed  $9X_3$ ). For each man we must total these two weighted scores with the constant term to get the weighted sum. With Adams, for instance, we total 84, 99, and 14, to get 197. These weighted sums then give the best statement of  $X_{11}$ , i.e., the best statement of the man's occupational proficiency, that it is possible to make on the basis of these two tests.

**Coefficient of multiple correlation.** The "coefficient of multiple correlation" is the correlation of the weighted sum of the tests with the criterion. That is, if all the original measures are reconsidered and each weighted according to the regression equation as is done in Table XXI, these weighted scores can then be correlated with the criterion to obtain the coefficient of multiple correlation. It can also be computed statistically from the partial coefficients without recurring to the original data and is often computed in both ways as a check on the work. (Cf. 490.) This coefficient of multiple correlation tells us just how valuable the tests are when combined in this manner and it is possible to see how much superior the combined weighted score is to the score in a single test. This procedure is also useful in determining the minimum number of tests that will give valuable results. If ten tests, weighted, give a multiple correlation with the criterion of .60, and four tests give a correlation of .58, it is probably unwise to retain the entire ten when four will do nearly as well and will occupy much less time in giving and scoring in the employment office. It can also be shown that the coefficient of multiple correlation is higher when the tests are weighted according to the regression equation than when they are weighted in any other manner.

Recurring to the previous example of tire finishers, it will be recalled that nine of the fifteen original tests were discarded on the basis of preliminary correlations. The remaining six tests (starred

in Table XX) were correlated by the products-moments method with the criterion and also with each other. For convenience, let us renumber the variables calling the criterion 1, the first starred test in Table XX, test 2, the next starred one test 3, etc. The correlations and intercorrelations are given in Table XXII. A

TABLE XXII. CORRELATIONS AND INTERCORRELATIONS OF BEST TESTS FOR TYRE FINISHERS

	1	2	3	4	5	6	7
1		.23	.51	.27	.49	.32	-.41
2			.36	.34	.32	.38	-.40
3				.66	.66	.70	-.24
4					.64	.66	.0
5						.71	-.22
6							-.01

coefficient in a given compartment of the table indicates the correlation between the variables at the left of the row and the top of the column that determine that compartment. For instance, the correlation between the criterion and test 2 is .23, that between test 2 and test 3 is .36.

The table gives some notion of the extent to which the various tests overlap. Test 4, for instance, appears to be measuring somewhat the same thing as tests 3, 5, and 6, because its correlations with these tests are respectively .66, .64, and .66, but it apparently has no relation whatever to test 7 because the correlation is 0. Test 7, on the other hand, seems unique because its correlations with most of the other tests are low. Its negative correlation with the criterion (-.41) is due to the fact that this is a test of reaction time and the quicker the reaction — i.e., the smaller the test score — the greater the efficiency.

With the correlation coefficients in the table it is now possible to compute the various partial correlation coefficients that are needed. A hint as to the method has been given above. After we have found the necessary coefficients like  $r_{12.3}$  and  $r_{12.4}$  we then com-

pute from these the coefficients like  $r_{12 \cdot 34}$  and  $r_{13 \cdot 24}$ . From these latter we go on to coefficients like  $r_{12 \cdot 345}$  till we finally reach partial correlations such as  $r_{12 \cdot 34567}$ .

We are then able to derive the regression equation by using these various partial coefficients. It is possible to weight all six tests and include them in a regression equation or to weight a smaller number. In the present instance the six tests were weighted and the coefficient of multiple correlation computed to determine how well these six weighted tests would predict ability at finishing tires. This coefficient proved to be approximately .63. It is to be noted that this is considerably better than the prediction that could be made with the best test taken alone, for the highest single correlation for a test with the criterion in Table XXII is .51 for test 3.

Investigation was then made as to how good a prediction could be made using only the three best tests, viz., 3, 5, and 7. When these were weighted according to a regression equation, the coefficient of multiple correlation proved to be .61. In other words, these three tests give almost as good a prediction of ability at finishing tires as do all six tests.

It is interesting to note with these three final tests their relative weight based on the partial correlations. In Table XXIII are given the original correlations of each of the three tests with the criterion and the corresponding correlations when the other two tests are kept constant. Test 3 correlates in the first instance to

TABLE XXIII. CORRELATION OF TESTS WITH ABILITY TO FINISH TIRES

	ORIGINAL CORRELATION	PARTIAL CORRELATION
Test 3.....	.51	.25
Test 5.....	.49	.23
Test 7.....	-.41	-.33

the extent of .51, but its partial correlation ( $r_{13 \cdot 57}$ ) is .25. Test 5 correlates originally to the extent of .49 and its partial coefficient is .23. Similarly, the partial correlation for test 7 is -.33. The partial coefficients are lower than the original. This is due to the

fact that the tests overlap, and when the overlapping is eliminated, the intrinsic relations are somewhat smaller. It is to be noted that, while test 7 has the lowest original correlation, it has the highest partial correlation. This is due to the fact that its correlations with the other tests are low, i.e., it is measuring a unique factor. Consequently it gets more weight in the final combination of tests.

In the actual process of deriving the regression equation, it is necessary to take account, not only of these partial correlations, but also of the average scores in each test and variability (standard deviation) of each set of scores. Consequently, the actual weights in the final equation do not look exactly like these partial correlations. The equation, however, does arrange it so that the relative contribution of each test to the total score is proportional to these partial correlations.

In the present instance the regression equation involving these three tests is:

$$X_1 = .02X_2 + .03X_3 - .014X_4 + 1.81$$

where  $X_1$  represents tire finishing;  $X_2$  indicates score in the test of cancelling adjacent pairs of numbers whose sum is 10;  $X_3$  is score in the test of finding consecutive numbers arranged irregularly in a square table; and  $X_4$  is visual reaction time. This equation is worked out in Appendix III in order to illustrate in more detail the derivation of such regression equations. The equation may be interpreted thus: if, for any given individual, scores in the three tests are available and his score in number 3 is multiplied by .02, his score in number 5 by .03, and his average reaction time by  $-.014$ , according to the equation, the sum of these values plus the constant 1.81 will give the most probable value of his ability in the job. It will not predict this latter with absolute certainty any more than a life insurance company can predict a person's date of death, but it does give the best estimate as to what a man can do in the job just as the insurance company determines the most probable age to which a person of certain status will live. The closeness with which such prediction for an individual case can be made depends on the correlation between the weighted tests and the

criterion, i.e., the coefficient of multiple correlation. If this correlation is quite high, it is possible to make a rather close prediction and the chance that the man will ultimately fall a long distance above or below the prediction is slight. If, however, the correlation is low there is a considerable chance that the prediction will fall wide of its mark.

The question of just how close a prediction can be made with correlations of different magnitude and of the chance that the employer must take in hiring employees with various test scores has already been discussed in Chapter VIII in connection with "Critical Scores." Furthermore, it can be shown statistically that a group of tests weighted according to the regression equation will correlate more highly with the criterion than will that same group of tests weighted in any other fashion, and hence this weighting process makes for more accurate prediction of vocational proficiency. It will be recalled that the above-mentioned discussion in the preceding chapter brought out the great desirability of raising the correlation from the standpoint of making fewer mistakes in prediction. Weighting the tests is one way in which the correlation can be increased. If a few days of statistical work will raise it from .50 to .60 and will double the efficiency of the methods in eliminating an undesirable class of applicants (as may often be the case — cf. p. 219), it is worth the effort and we find herein the justification of the rather laborious task of weighting the tests.

The method of devising a group of tests to measure the mental components of the occupation has been rather widely used. Some of the work has been done with the statistical refinement above described, but frequently rougher methods have been employed. In most instances, however, there has been some effort to compare efficiency in the tests with efficiency in the job. A few other cases of the former sort will be described further to illustrate the method.

#### EXAMPLES WITH PARTIAL CORRELATION TECHNIQUE

**Telegraphers.** A group of drafted men in a telegraph school were studied with a view to measuring potential telegraphic aptitude. (608.) The criterion consisted of their efficiency in receiving after 100 hours practice. A considerable number of tests



that seemed to measure factors involved in the occupation were selected. A rhythm test consisted of writing dots and dashes in a rhythm that was presented to the subject in auditory fashion. There was a conventional opposites test and an analogies test (cf. examples 16 and 17, Chapter IV). There was also a test of following directions (cf. example 32, *supra*), a completion test which involved filling in missing words in a paragraph (cf. example 31, *supra*), as well as tests of spelling and arithmetic. The original correlations for the best tests appear in Table XXIV. The remaining tests were discarded on the basis of preliminary correlation with the criterion. The partial correlations of each test with the criterion, keeping all the other tests constant, likewise appear in the table. It will be seen from these figures that the rhythm test is the most important of the group. Its original correlation is the largest, and then, when all the other tests are kept constant and the intrinsic relation of test to criterion determined, this test shows a still greater superiority to the others. The regression equation was worked out in the method above described to determine the best possible weighting of the tests. The coefficient of multiple correlation between the weighted sum of the tests and the criterion is .53. The rhythm test alone correlates to the extent of .48 so the inclusion of the other tests does not very greatly increase the predictive value. The rhythm test manifestly carries most of the load.

TABLE XXIV. CORRELATIONS OF TESTS WITH EFFICIENCY  
IN TELEGRAPHY<sup>1</sup>

	ORIGINAL CORRELATIONS	PARTIAL CORRELATIONS
Rhythm test.....	.48	.31
Opposites.....	.42	.21
Analogies.....	.29	.02
Directions.....	.27	.06
Completion.....	.21	.09

<sup>1</sup> After Thurstone.

**Aviators.** A set of tests was developed to predict aptitude for flying an airplane. A preliminary selection of a considerable number of tests was made: various kinds of simple and choice reaction times, ability to detect sudden changes of equilibrium, ability to note slight and gradual changes of equilibrium, susceptibility to fatigue, estimation of distances and velocities, ability to detect slight differences in sounds, speed of tapping, emotional stability (steadiness of pulse and breathing and hand tremor after a revolver shot), and a number of other measures. Some of the tests were eliminated after giving them to cadets in a number of ground schools and comparing tests with subsequent flying ability. The remaining group of tests was then tried out more fully at flying schools and scores correlated with instructors' estimates as to flying ability and with the number of hours of flying instruction with double control before the cadet was allowed to fly the ship alone. The measures finally used consisted of a brief mental alertness test, evaluation of certain items in the personal history blank especially with reference to athletic activities, the extent of swaying as the subject stood at attention with a pointer attached to the top of his head writing on smoked paper, the angle through which the subject's chair could be tilted very slowly before he was aware of the direction of the tilt, changes in breathing and hand tremor after a revolver shot, choice reaction to a sudden change of equilibrium as the platform on which the subject was seated tilted suddenly to left or right, and auditory plus visual reaction time minus equilibrium reaction time. This last item is of particular interest, for it developed that not only was equilibrium reaction time significant, but that time relative to the other reaction times was especially so. That is, a person whose equilibrium reaction time is quick relative to his other reaction times (visual and auditory) is more promising as an aviator. This differential score was discovered by partial correlation and would probably not otherwise have been found. These tests were then all weighted by the regression equation procedure and the coefficient of multiple correlation between the weighted sum of tests and the criterion was approximately .70. This final result may be shown in another way with the data for a group of cadets at Kelley Field. In Table XXV the men are

grouped into eight classes on the basis of their weighted test score (first column). They were estimated by their instructors as good (g) or poor (p). Each letter in the second column indicates an individual aviator. The preponderance of poor men among the

TABLE XXV. COMPARISON OF WEIGHTED TEST SCORES WITH FLYING ABILITY

WEIGHTED SCORE	INSTRUCTORS' ESTIMATES	AVERAGE DUAL TIME
-75 or worse	pppppp	15.5
-50 to -74	ppppppg	11.2
-25 to -49	pppppppppp pppg	10.2
0 to -24	pppppppppp ppppppgggg ggggg	10.0
0 to +24	ppggggg	9.2
+25 to +49	ppgggggggg gggg	9.2
+50 to +74	ggggggg	7.5
+75 or better	g	7.2

low scores and the preponderance of good men among the high scores is obvious. The last column gives the average dual time — i.e., average number of hours spent flying with an instructor and dual control before being allowed to fly alone — for each of the eight groups. Whereas those with lowest test scores required fifteen hours of flying instruction on the average the best ones in test score required only about seven hours and with decreasing test scores there is a consistent increase in dual time.

**Clerical workers.** A set of ten different tests was given to a group of clerical workers and preliminary correlations made with managers' estimates as to ability in general clerical work. (102.) It was obvious from the preliminary data that four of the tests might be discarded at once. The remaining six were intercorrelated and determination made of the final multiple correlation for weighted

test scores and criterion. It developed that four tests were practically as good as the six. Test 2 involved underlining adjacent letters that formed words in a printed page of unspaced letters. Test 3 involved going down a column of numbers and adding 17 to each. Test 4 was an "analogies" test in multiple choice form (example 17, Chapter IV). Test 5 consisted of a test with every pair of adjacent letters in each word separated by an irrelevant letter, as:

atzhxicsv binam lak jmhgnftdasla qtweesrtt  
 this is a mental

The subject was to decipher the test by reading every other letter in each word. The original correlations of test and criterion and those same correlations when the other tests are kept constant by partial correlation are given in Table XXVI. The tests manifestly

TABLE XXVI. CORRELATION OF TESTS WITH GENERAL CLERICAL ABILITY

	ORIGINAL CORRELATION	PARTIAL CORRELATION
Test 2.....	.40	.14
Test 3.....	.34	.17
Test 4.....	.31	.18
Test 5.....	.47	.23

overlap somewhat, as is shown by the fact that the partial correlations are considerably lower than the original. The regression equation for these four tests is as follows:

$$X_1 = .019 X_2 + .011 X_3 + .088 X_4 + .015 X_5 - 251$$

The multiple correlation between the weighted test scores and the criterion is .56, which it will be seen is appreciably better than that obtained with any single test.

EXAMPLES WITH LESS STATISTICAL REFINEMENT

Telephone operators. The foregoing are typical of tests for the mental components of the job evaluated by the aid of partial corre-

lation procedure. The majority of such experiments, however, have in the past been performed without this refinement, and while further study and weighting of the various tests would doubtless have improved matters somewhat, the tests have nevertheless been fairly successful in predicting vocational aptitude. A few such studies will be described by way of illustration.

One of the earliest of these was a test for telephone operators. (408.) These tests were given to girls in a telephone school. They comprised memory span for numbers — i.e., the maximum number of digits that could be repeated after a single reading; a cancellation test, crossing out certain letters on a page; a memory test by the method of word pairs; a test of card sorting; a motor coördination test — tapping rapidly in sequence three crosses on the blank; and speed of association reaction to a stimulus word. The scores of the girls were ranked in each test and then the average rank computed for each girl. These average ranks were then compared with proficiency in the telephone school after three months' service. There was a marked although not universal tendency for those with the better test scores to be more proficient in actual service. The company, moreover, had surreptitiously introduced among the supposed pupils in the school a number of expert operators. These experts made high scores.

**Insurance salesmen.** A group of insurance salesmen was given a form of the will temperament test (cf. example 38, Chapter IV); a test for objections to purchase in which hypothetical objections raised by a purchaser must have an appropriate answer given for them; a questionnaire designed to indicate a person's interests, another questionnaire containing items of personal history, a brief intelligence test and a test of "social relations" — i.e., information regarding music, games, sports, religion, theatrical matters, etc. (468.) These test scores were compared with the criterion and the most useful determined. They were then weighted as follows (the items the names of which do not obviously connote the nature of the test being part of the will-temperament test): flexibility in disguise 7, personal history 6, speed of decision on personal traits 6, objections to purchase 5, freedom from self-consciousness 5, care for detail 5, interest analysis 4, speed of decision 3, intelligence 2,

social relations 2, speed of movement 2. When the scores are weighted in this fashion and plotted against the criterion, the scatter plot appears encouraging and shows a fairly close relation.

A notion of the predictive value of this composite set of measurements may be obtained from Table XXVII. For instance, of the

TABLE XXVII. SUCCESS OF INSURANCE SALESMEN COMPARED WITH COMPOSITE TEST SCORES <sup>1</sup>

COMPOSITE SCORE	SUCCESSFUL SALESMEN	DOUBTFUL SALESMEN	UNSUCCESSFUL SALESMEN
3 to 22	92	8	0
0 to 2	44	28	28
-1 to -15	0	33	67

<sup>1</sup> After Ream.

men who scored between 3 and 22, 92 per cent were successful, 8 per cent doubtful, and none unsuccessful. Of those who scored between 0 and 2, 44 per cent were successful, 28 per cent doubtful, and 28 per cent unsuccessful. This shows a fair tendency for those with high composite scores to be more successful than those with low scores.

#### MISCELLANEOUS BRIEF EXAMPLES

**Office workers.** The majority of mental test projects have employed the foregoing technique of measuring separately the various mental factors involved in the job. To give a notion of the scope of this work, very brief mention will be made of a considerable number of different vocational fields in which this technique has been applied. Many such studies have been made with various kinds of office workers. A few tests were given to good typists who typed more than 540 sheets per day and to poor typists who typed less than 430 sheets. (263.) In a test of giving opposites, the former group takes 70 seconds on the average while the latter takes 78 seconds; the good typists require 60 seconds in a color-naming test and the poor typists 65 seconds; in a test of substituting symbols for numbers the good ones make 2 errors on the

average and the poor ones 12. A study was made of champion typists and their proficiency in certain motor tests compared with the proficiency of persons in general. (71.) The tests comprised motion of the forefinger, of the hand bending at the wrist, of the forearm bending at the elbow, and of the upper arm bending at the shoulder. The standard proficiency of individuals of various ages had been previously established in another connection. The tests were given to contestants in a national typewriting contest and also to a number of ex-world champions. The superiority of each individual to the standard established for his age was noted. The present world champions are approximately 30 per cent superior to this standard; the ex-world champions 26 per cent, the amateurs in the contest 15 per cent, and the world school champions 13 per cent superior to the standard. Studies of typewriting have also been made with correlation technique. In one instance upwards of twenty different tests were tried on many different groups and test scores correlated with production or with teachers' estimates. (415.) The correlations varied considerably from one group of typists to another, sometimes being as high as .70. The best tests on the whole involved immediate memory span for sentences, ability to follow complicated directions, finding the products of pairs of numbers in a table in which it was necessary to locate the appropriate row and column in order to find the product, a completion test, i.e., filling in omitted words (cf. example 31, Chapter IV), and a spelling test. In a similar study with a single group of subjects, four tests proved to correlate appreciably with efficiency in typing. (620.) The tests with their correlations are as follows: tapping on a typewriter key .54, underlining the letters *zn* wherever they occurred together in a page of letters .68, a substitution test (cf. example 14, Chapter IV) .52, and underlining pairs of adjacent numbers whose sum is 9 in a page of random numbers .41. A similar study of stenographers yields correlations a little smaller than the preceding. (488.) For hard "directions" (cf. example 33, Chapter IV) the correlation is .46, for giving opposites .45, and for substitution .40.

Studies of office workers have not been confined to typists and stenographers. Billers in a freight department were given tests of

following directions, completion and substitution. (230.) The agreement between their test performance and vocational proficiency is "very marked." In another study of operators of a billing machine the correlations with the criterion are as follows: intelligence .48, cube imitation (cf. example 43, Chapter IV) .45, cancelling A's in a page of random letters .37 and sorting cards .39. (308.) With Hollerith machine (a statistical machine) operators, opposites, substitution, completion, and similar tests give rather small correlations, but when weighted according to a regression equation five such tests yield a coefficient of .45. (361.)

**Factory workers.** Some industrial concerns have adopted various series of tests which they use without any very consistent effort to determine their validity. Typical of such programs are the tests used in an electrical manufacturing concern. (236, 427.) One test involves assembling a cube that is cut into nine irregular pieces; another comprises a metal block containing one hundred holes and the subject inserts pegs in these holes three at a time. Nothing is said regarding the use or validation of the tests. A similar concern is using a much more elaborate series of tests. (496.) Visual acuity is measured with small wires of different size; space perception is tested by bisecting lines, locating the middle of circles and dividing arcs into equal portions; "distributed attention" is measured by sorting tags and at the same time turning two miniature hour-glasses as soon as they require it; motor control is determined by holding a stylus in small holes without touching the edge or by adjusting a surface on which a ball can roll. Distribution curves of test scores made by various groups of workers are given, but no actual correlations with vocational proficiency. In a camera concern a mechanical test involving fitting small metal pins into holes in a board is said to have been useful. (355.)

In most of the factories that employ tests, however, some effort has been made to validate the tests. A series of tests was devised for cabinet workers. (20.) One involved lining up a straight edge with a target and the error was measured by a micrometer device; another consisted of the handle of a plane with a device to record the length of stroke as the subject tried to move his handle a fixed distance at constant rate; another device indicated the constancy



of pressure on the handle of the plane; another test involved judging cross-sections of objects such as cylinders taken in various directions; another consisted of drawing angles and another of adjusting two surfaces with a micrometer screw so that they felt flush when tested with the finger-tip. The workers were grouped into three classes on the basis of their actual proficiency in cabinet work. The correlations of the first four tests above mentioned with this criterion are respectively .87, .83, .75, and .81. The correlations of these tests with each other are much lower than this, which is especially favorable if it is desired to combine them into a regression equation.

In the sorting department of a feather factory three tests proved rather satisfactory with a small group of workers. (99.) One test involved the use of pieces of cardboard of somewhat the shape of an ostrich feather. These cards were shown for 10 seconds each and the subject judged their length and width. The correlations of ranks in this test with foremen's ranks as to proficiency in sorting feathers is .65. In a color discrimination test the correlation is .58 and in a brief actual sample of sorting feathers it is .76.

In a chocolate factory fair correlations with the criterion (three classes of ability in packing chocolates) are reported, although the tests are not described. (171.) The correlations are as follows: hand and eye coordination .59, speed test with a peg board .49, speed test with beads .44, visual memory .44, substitution .40, and perception of equal distances .37.

Tests of a small group of weavers showed rather high correlations with the manager's ratings of the workers. (548.) Dexterity in twisting things with the fingers was tested by having the subject turn a standard thumbscrew. The average length of each turn correlates with the criterion to the extent of .73 and the speed of turning to the extent of .63. A test of disparate attention, in which cards were shown and the subject noted simultaneously three different aspects of the card, correlates to the extent of .51. A pattern discrimination test, in which patterns were presented in pairs and the subject noted the difference between those of each pair, gives a correlation of .49. A similar test, in which two patterns were shown successively rather than simultaneously, gives a correlation

of .38. Three other tests each give correlations of .41. One involved a series of spools of yarn of varying thickness which were to be arranged in order of magnitude. Another involved threading a stiff wire through eyelets suspended in a frame. The last involved planning ahead as typified by performance with a maze. When the tests are combined into a single score, the correlation with the criterion is .77.

In a printing establishment a series of tests was adopted for typesetters. (342.) The first consisted of spelling and punctuation. The second involved reading a badly written paragraph with many ink blots in it (something like a completion test). The third involved setting up a sentence that used only nine different letters with nine boxes of type provided. The fourth consisted of copying a long sentence on a typewriter. In these last two tests the number of times the subject looked at the copy was noted. The foregoing tests were used for employment purposes. Standards were, however, determined arbitrarily rather than by correlation procedure. A more scientific study of typesetting was made later. (412.) Six tests were evaluated by comparison with efficiency in the job and four were retained. A cancellation test — cancelling every letter *e* in a page of nonsense type — gives a correlation of .64; a substitution test — substituting digits for letters — gives a correlation of .58; a directions test (cf. example 33, Chapter IV) has a correlation of .57, and a test of inserting pegs in holes gives a correlation of .57. With a group of typesetters in another company, the correlations for these same tests are respectively .59, .50, .60, and .56, showing pretty fair agreement between the two groups. When the tests are weighted approximately according to a regression equation, the correlation for the group of tests in the first company is .71 and in the other .80.

Telephone and telegraph operators. Following the lead of earlier work cited above, various studies have been made of telephone operators. In one case the first test made use of a window at which six letters might appear. (294.) The subject reacted to certain of them with four telegraph keys and his reaction time was noted. In a second test the subject reacted with an actual telephone plug, putting it into certain jacks according to the symbol

that was presented. A third test involved a series of patterns which were presented briefly and which the subject attempted to reproduce from memory. A final test involved memory span for numbers and also logical memory. These tests were evaluated in a preliminary way and ranked in the order of their importance for prognosis, although no correlation procedure was used. Somewhat similar statistical technique was used in another study of telephone operators. (199.) It involved a completion test, a test in which symbols were reproduced after a short exposure, a "monotony test" in which the subject closed two electric contacts alternately and counted aloud, various memory tests, steadiness of the hand, auditory acuity, and a sort of choice reaction to letters that moved across the field. A consideration of some of the extremely good and poor operators showed some correspondence between tests and criterion. A similar study was made with greater statistical refinement. (170.) The girls in a telephone exchange were ranked by the manager and the following correlations with test scores obtained: auditory memory span for digits .47, imagining a paper folded so as to have the creases bound several sections and then touching the middle of each of these sections successively in time with a metronome .44, sorting cards into four boxes .43, serial memory — names of ten cities read aloud in succession — .42, variability in simple visual reaction time .32. All the other tests give correlations less than .30. If the tests are weighted equally, the combined scores correlate with the manager's estimate to the extent of .54. If only the best tests are used in this manner, the correlation is .70. The development of prognostic tests for telegraphers has already been cited. An earlier effort along the same lines yielded a group of tests that gave a correlation of about .50. (265.) In another study the tests employed were more similar to actual telegraph practice so that they comprised a mixture of trade tests and innate capacity tests. One test involved receiving actual telegraph code. Another necessitated receiving words and selecting the first letter of each. These first letters made a sentence. Dots and dashes (nonsense) were received in groups of 3 to 10. Ability to discriminate differences in the pitch of a tone was measured. Critical scores in these tests were established without, however, evaluating them statistically.

**Motormen.** In addition to the tests of total mental situation for motormen described above, there have been attempts to measure the mental components of the occupation. One of these was quite elaborate. (210, 617, 618.) The motormen were tested with reference to color blindness and visual acuity. Vision in dim light was determined by placing the subject's head in a box and having him discriminate letters displayed therein. A steadiness test similar to example 1, Chapter IV, was given with the exception that both hands were used simultaneously and held the stylus in the manner in which controls on a trolley are ordinarily held. In a similar test a rod was moved through an arc like that described by the controller on a car without touching the edges of the slot. Two vertical rods some ten or twelve feet long were balanced and the subject required to catch either as soon as it began to tip. A scale at the top of the rod indicated how far it tipped before the subject noted it. Knowledge of traffic rules was tested with miniature cars on a street intersection. Certain ones were suddenly lighted and the subject required to tell which had the right of way or which should start first. A test with a miniature track and lights flashing on to indicate pedestrians and vehicles, similar to that described above as a test for total situation, was included in this series. To measure emotional stability the subject was instructed as to what levers he should operate in case anything unforeseen happened. Then it was arranged so that the floor beneath him gave way slightly or so that some wires a short distance in front of his face short-circuited and created a substantial arc. To get at the subject's mechanical insight he was shown pictures or models of various arrangements of gears and required to state in what direction a certain member would move if another designated one moved in the direction of the arrow. The tests apparently were not correlated with a criterion, but were actually used for hiring employees. Some twenty-five per cent of the applicants were rejected on the basis of these tests. In the first year the men who had been hired without being tested had fifty per cent more accidents than those hired on the basis of the tests. The training time for these latter was shortened by about 120 hours. A briefer set of tests along these same lines was said to have shown some

agreement with the men's service record. (198.) In another similar project the subject reacts to the presence or absence of certain lights with two hand levers and two foot pedals — using about the same muscles that he used in actually operating a car. (635.) The average motorman makes six mistakes in the test. Of those classed by their supervisors as "fairly safe," 75 per cent make eight or more mistakes, while of those rated as "safe" 35 per cent make eight or more mistakes, and all of those rated as "very safe" make less than six mistakes. Another test which showed some relation to ratings on courtesy involved items of judgment like the following:

1. If a passenger bawls you out when you do not deserve it, what would you do?
  - .... Call the conductor to help put him off the car.
  - .... Shout back at him.
  - .... Say nothing to him, but report the incident to your superintendent.
  - .... Explain quietly to the man that he is wrong.
  
2. If an intoxicated man was annoying passengers in your car, what would you do?
  - .... Put him off the car.
  - .... Pay no attention to him.
  - .... Turn him over to the nearest officer.
  - .... Report to the train dispatcher.
  
3. If a passenger asks you to speed up your car so that he can catch the car ahead, what would you do?
  - .... Speed up to satisfy the passenger.
  - .... Tell the passenger to ask the conductor.
  - .... Decrease the speed of your car.
  - .... Refuse to comply with the request of the passenger.

**Automobile drivers.** An approach has been made to the problem of the mental qualifications necessary for safe operation of an automobile. (402.) One such study was confined to the reaction time of the driver under actual conditions of operating the car. A tachometer (a device more accurate than the usual speedometer) was arranged so that when the car reached a certain speed a pistol on the running board was automatically fired. This pistol was

loaded with a red pigment which made a mark on the pavement. The subject was to apply the brakes the instant he heard this signal. The application of the brakes fired a similar pistol. The distance between the two marks on the pavement made it possible, knowing the speed of the car, to compute the reaction time. Taxi-drivers were quicker in this test on the average than were other subjects. Some of the best reacted in as short a time as 0.3 second. At 20 miles an hour a reaction time of 0.5 second will begin to stop the car in 15 feet, while a reaction time of 1.5 seconds will allow the car to go 45 feet before beginning to stop. It is, of course, possible to measure a person's reaction time in the laboratory before he ever drives an automobile at all, and while experiment only would settle the matter it is probable that persons who reacted quickly in the laboratory would do likewise under actual driving conditions.

A more elaborate study of drivers has been undertaken by the Yellow Cab Company. (544, 545, 546.) In Chicago some 21 per cent of the applicants for employment have been rejected on the basis of a poor showing in certain tests and as a result there has been a substantial decrease in accidents. According to a recent statement this decrease amounted to 34 per cent. A brief description of the tests follows. An abbreviated intelligence test something like the Army test is useful in eliminating applicants of extremely low capacity. With those of average or superior intelligence, however, the test does not appear differential of ability in the job.

Reaction time during fear is determined with a rather complicated equipment. The subject sits at a table in a dark room. He closes two switches by pressing on pedals with his feet and also closes another with his left hand. He can then light any one of a number of small lamps on the board before him by taking a plug in his right hand and touching the proper terminals on the board. He is told to operate the small switchboard so as to light the lamps in order. At his right are two other switches, one operated by hand and the other by foot. The subject is told that if anything unusual happens while he is lighting the lamps, such as a flash of light or an electric shock, he is to shut off the current by operating these two

switches at his right. The test is taken in almost complete darkness. There is a board mounted about ten inches in front of the subject's face containing spark gaps. During the test the subject is without warning given an electric shock and an arc created in the spark gap in front of his face. The time taken for him to make the appropriate reactions with the two switches at his right is automatically recorded. In another test the subject is given a board on which are cut four irregular lanes. They vary in length, but have the same terminus. The shortest is the narrowest and the longest is the widest. There are various irregularities and sharp turns in these lanes. The subject traces through the lane with a stylus avoiding contact with the edges. Every such contact is recorded automatically and another recording device makes it possible to note the speed with which the subject moves the stylus at various portions of the course. The subject is scored according to whether he selects the lane that will enable him to make the best time and whether he slows down at points of difficulty such as corners. Other tests involve tapping on a telegraph key and maintaining a constant pressure against a spring. Reaction time is also measured with foot pedals and hand keys. A final test involves a number of toy vehicles whose motion is controlled electrically. They are moved in various directions and at various speeds and the subject is required to indicate the point at which they will pass or overtake each other. This work is at present still in the experimental stage, but its use thus far has decreased appreciably the number of accidents.

**Public service.** A number of occupations in the public service have been investigated in the same manner as have those in the commercial field. Tests for fire-fighters have been proposed. One such proposal involves reaction time measurements and measures of emotional stability through a record of breathing or tremor when a sudden stimulus such as a revolver shot is applied. (15.) Another proposal presents a number of tests each to receive the weight indicated in parentheses: Army intelligence test (2); observation — looking at a picture of a fire scene and then answering questions about it (1); understanding printed material dealing with fire-fighting — the subject is handed the material and answers ques-

tions by consulting it (1); memory for verbal orders of the sort issued during a fire (1); education and experience (1); medical and physical examination, strength and agility (3). (466.)

In quite similar fashion tests have been suggested for policemen with the weights indicated: Army intelligence test (2); accuracy of observation — questions on a picture shown previously and also recording auto tags from memory (1); memory tests — recording facts from a description that is read and identifying photographs shown once and then mixed with other photographs (1); understanding laws and police rules — the subject is given a copy and answers questions by referring to the copy (1); police duties — identifying crimes from definitions and descriptions of cases (1); education and experience (1); personal traits as determined by an interview (1); medical and physical tests (2). (584, 600.)

Similar qualifications have been suggested for a government hospital attendant. (400.) The items involve a written duties test (understanding hospital rules, recognizing certain mental diseases and general information), education, personal traits manifest in an interview, medical and physical examination. Similarly, for a janitor there have been suggested a mechanical intelligence test (example 45, Chapter IV), tests dealing with work and materials and understanding printed material dealing with cleaning and maintenance. (583.) Other tests along these lines are being developed by the Bureau of Public Personnel Administration. Some of them, of course, border on trade tests (*infra*) at least in certain portions, but innate factors such as intelligence are involved in most of them. They include tests for prison guard (586), food inspector (403), bacteriologist (256), library assistant (94), pathologist (255), playground supervisor (92), road inspector, (535), shift engineman (536).

Miscellaneous. Hydrophone listeners were selected on the basis of tests of keenness of hearing, accuracy of discriminating sounds, memory for pitch, rhythm, and quality of a sound, pitch discrimination, discrimination of rhythms and sound qualities, general accuracy and ability to follow directions. (418.) The first group selected in this way and sent into service was reported as "far and away the best" group that had hitherto been received.



In an attempt to devise methods for predicting aptitude for public speaking students in a class read a selection from the rear of the room and were judged by their fellow students. (653.) This criterion was compared with results with the phonograph records for the Seashore tests of musical talent (cf. example 8, Chapter IV). The correlations are as follows; sense of pitch .48, tonal memory .35, sense of time .19, and sense of intensity .18. Obviously the test for pitch carries most of the load and if the tests are weighted according to a regression equation the correlation is only .49.

Tests have even been suggested for judges in criminal courts. (393.) The principal one involved observation of a picture or a standard series of events and then questions regarding what had been observed. Others comprised memory for physiognomies and a completion test. Marked individual differences were found between different judges but no criterion for evaluating the tests was available.

A series of tests has been devised for predicting journalistic aptitude. (181.) The first involves the subject's ability to discriminate the relative importance of items, i.e., his "nose for news." It is a story of a fire and the subject marks the most important of various groups of items. The next presents various situations which might confront a reporter. Each situation has listed with it several alternative courses of action which the reporter might take. The subject in each instance has to check the best alternative. The third test necessitates the checking of correct definitions in a list which comprises correct and incorrect ones. The next involves checking errors in an exciting story. The next is a conventional opposites test and the last presents a picture of an accident comprising a lot of detail and after a brief observation questions regarding items in the picture are to be answered. This test was given to students in journalism classes in a number of universities and correlated with instructors' estimates as to journalistic ability. The correlations vary with the different groups but average around .40.

#### FOLLOW-UP PROCEDURE

After a psychologist has developed a test or series of tests for predicting aptitude in a certain occupation his task is not completed

as far as that occupation is concerned. The methods can be put into use in the employment office for selecting workers. It is desirable, however, to keep a record of their scores in the tests taken at the time they are hired and subsequently to compare their test scores with their ability in the job after they have learned it. After a sufficient number of applicants have taken the tests and have been at the job long enough to reach their maximum efficiency it is well to get figures as to the latter ability in much the same manner that the criterion was originally determined and then to compare the original test scores with this new criterion. This will serve to vindicate the whole procedure, for while it is probable that tests devised originally to differentiate the good from the poor employees will serve likewise in differentiating the good from the poor applicants it is well eventually to determine empirically if this is the case. Furthermore, an occasional check on the value of the methods is desirable because there may sometimes be changes in the general type of applicants, the methods of training or even the methods of work that will render the original tests invalid. This follow-up procedure has a further advantage in that it may be possible from time to time to introduce slight changes in method. It may be desirable to give one or two tests in addition to those originally standardized and evaluate these subsequently with a view to including them ultimately in the regression equation — possibly replacing some of the original ones.

It is well for the employment psychologist to keep in touch with his original work. It is, of course, often necessary to develop methods, make them as objective and fool-proof as possible and then turn them over to untrained persons in the employment office for routine administration. This is probably not ideal. The technique of mental examination is more reliable in the hands of a person with psychological training. Unforeseen contingencies may arise. Very frequently extraneous reactions which the applicant makes, quite apart from his actual test performance, are of vocational significance and only the trained examiner can make the most of this "clinical picture." The time will come when large concerns will have a psychologist permanently attached to the staff — just as they now have chemists, physicists, or engineers —

to maintain constant supervision over the personnel and other work that is psychological in character. An industrial concern is, in a way, a psychological laboratory in which the problems are not all solved and the methods devised for all time but in which research may well be continuously in progress.

#### SUMMARY

In devising a set of tests for the mental components of the occupation the preliminary procedure of analysis is similar to that for the test of total situation. It is then necessary to select and devise tests for the various components that the analysis reveals. No test measures, of course, an isolated mental factor but this procedure will probably bring better results than selecting tests at random. The more tests used the greater the chances of finding some which have high correlations with the criterion. The number evaluated generally depends on the length of time for which the subjects are available.

The tests selected must be given to subjects and evaluated to determine which to retain and which to discard. Usually some rough correlation technique is adequate to eliminate the worst tests. The remaining ones are then correlated more carefully with the criterion and with each other in order to assign each test its proper "weight" in the total score. It is not desirable to weight each test according to its correlation with the criterion because some of the tests may be measuring substantially the same factor while others may involve more independent factors. This overlapping may be ascertained by correlating the tests with each other.

The technique of partial correlation makes it possible to eliminate the effects of this overlapping. By this technique one computes what the correlation of a test and the criterion would be if based on subjects who all had the same ability in the other tests. This shows the intrinsic relation of a test to the criterion and affords a more adequate weighting for each test than does its original correlation which takes no account of the overlapping. A consideration of the partial correlation formulæ shows that the best test for the present purpose is one which has a high correlation with the crite-

tion and a low correlation with the other tests, for this will tend to make its partial correlation with the criterion high.

A regression equation can then be derived which expresses probable vocational aptitude in terms of the tests. It indicates the weight or constant number by which to multiply each test score so that the weighted sum will give the best possible prediction of the criterion. The weight for a test is roughly proportional to its partial correlation with the criterion with the other tests kept constant.

The coefficient of multiple correlation is simply the correlation of the weighted sum of the tests with the criterion. This indicates how valuable the tests are for the purpose in hand and shows how much the weighted sum of the tests is superior to any single test.

Various examples of tests for the mental components of the job were given. Some of these employed partial correlation technique; for instance, tests of attention and reaction time for tire finishers, tests of rhythm and association for telegraphers, tests of sense of equilibrium, reaction time and emotional stability for aviators and various tests of attention or association for clerical workers. In such cases the tests were weighted according to a regression equation. In other examples less statistical refinement was employed although the work was analyzed into its components and effort made to measure these separately. Sometimes the tests were weighted equally or sometimes an arbitrary weighting adopted. Some of the examples were tests of coordination, memory and association for telephone operators, tests of temperament for salesmen and tests involving classification of items and noting similarities and differences for filing clerks. The method of testing the mental components of the job has been rather widely used and to illustrate its scope brief examples were given of such tests for office workers, industrial operatives, motormen, taxi drivers, telegraph and telephone operators, workers in the public service such as firemen and policemen and miscellaneous persons such as public speakers and journalists.

When a test project has been developed and put into practical use it is desirable to follow up the results for a time and see whether the new employees hired on the basis of the tests actually conform

to the prediction. This will serve as a subsequent validation of the whole method. It also makes possible minor revisions of the tests. If the psychologist is able to keep in touch with the work it is possible to have a continuous program of occasional addition and revision with a view to gradually increasing the validity of the employment methods.

## CHAPTER X

### INTELLIGENCE AND VOCATIONAL APTITUDE

THE preceding chapter discussed tests for special mental capacity in so far as they may be used to predict occupational success. Perhaps most of the employment problems with which a psychologist deals are of this type. Occupational misfits are usually lacking in some of these special respects. The foreman will state that the worker "can't remember numbers," or "does not put his mind on his work," or is "too slow." With modern industrial organization the majority of jobs necessitate the acquisition of a relatively small number of habits; and it is a question of whether the applicant has the special capacities such as memory, attention, or quick reaction time, that will facilitate the formation of those habits. There are other cases, however, in which the job apparently does not call for such specialized mental equipment, but rather for an all-round ability, a general mental alertness, or a facility in adapting one's self to a new situation. Any employment man will tell you about the worker who does not seem to fit anywhere, who is "stupid," who has to be told repeatedly what to do, and who does not "use his head" if anything unusual occurs. The trait involved here has usually been termed intelligence, and various tests have been devised to measure it. As stated in Chapter IV, it does not matter whether this general ability is called intelligence or something else, and its exact nature is of little consequence. If the results of these general tests enable us to predict occupational success, that is all that is required. The present chapter will be devoted to the use of intelligence tests for predicting vocational aptitude.

#### OCCUPATIONAL HIERARCHY

Occupational studies in the army. One question that arises in connection with such tests in employment psychology is whether a certain minimum of intelligence is required for different occupations. It seems plausible that a person will in the long run get

about as high in the occupational scale as his intelligence warrants, and that, if we determine the average intelligence of persons in a certain occupation, this will tell us something about the general ability required for that occupation. Data bearing on this point were available as a result of giving the army intelligence test to a large number of drafted men. In connection with their examination a record was made of their previous occupation. It was a relatively simple matter then to select a group of laborers, or a group of machinists, or a group of professional men, and compute the average intelligence of each occupational group. The results are rather illuminating (419, 819) and a typical portion of them is shown in Table XXVIII.

The scores are in terms of the actual number of points made out of a possible 212. The first column gives the first quartile score, i.e., the score which one fourth of the group fails to surpass. This is the same as the 25 percentile (cf. p. 97). The next column gives the average. The third one gives the third quartile, i.e., the score which three fourths of the group fails to surpass. This is the same as the 75 percentile. Putting it in another way, if the men in a given group were lined up in order of intelligence, with the lowest at one end and the highest at the other, a point one fourth of the distance up from the lowest end of the line would be the first quartile, the average would be approximately at the center of the line, while the third quartile would be three fourths of the distance up from the lowest end. The distance between the first and third quartiles obviously includes the middle half of the group, and it is often used as a rough measure of the variability or scatter. If the first and third quartiles are close together, this indicates that the individuals are "bunched" or have a small variability, while if these quartiles are widely separated, it shows that the individuals are scattered or have a large variability. The army test also utilized letter grades, C being average intelligence, B high average, and A superior intelligence. The last column of the table gives the per cent of each group in class A or B.

The table gives only a few of the many occupations that were computed in this fashion, but it is sufficient to afford a notion of the general trend. There is rather definite evidence of an occupational

## INTELLIGENCE AND VOCATIONAL APTITUDE 267

TABLE XXVIII. INTELLIGENCE OF OCCUPATIONAL GROUPS

	FIRST QUARTILE	AVERAGE	THIRD QUARTILE	PER CENT IN CLASS A OR B
Engineer officer.....	144	162	176	96
Medical officer.....	117	129	152	77
Civil engineer.....	99	117	143	68
Accountant.....	98	117	136	68
Stenographer or typist.....	93	115	138	62
Mechanical draftsman.....	84	114	139	59
Mechanical engineer.....	73	110	137	47
Bookkeeper.....	77	101	127	46
Filing clerk.....	74	97	126	40
General clerk.....	74	96	121	40
Railroad clerk.....	69	91	115	37
Telegrapher.....	61	85	110	28
Telephone operator.....	57	70	109	20
Auto assembler.....	51	68	97	18
General mechanic.....	48	68	94	14
Toolroom expert.....	50	67	92	9
General auto repairman.....	43	65	91	13
Telephone lineman.....	43	64	88	12
General carpenter.....	40	60	84	9
Baker.....	40	59	87	11
Bricklayer.....	37	58	88	11
Truck-driver.....	37	58	83	11
Barber.....	34	55	78	7
Boiler-maker.....	31	51	74	9
Teamster.....	30	50	72	6
Miner.....	40	49	71	5
Farmer.....	30	48	73	7
Laborer.....	28	47	68	4

hierarchy. At the bottom of the intelligence scale we find the unskilled laborers; higher up we find those in more skilled mechanical occupations; above those are the clerical and business workers; and at the top those in the professions. The results should perhaps be somewhat qualified in view of the fact that some of the better tradesmen were exempt from the draft, and of the further fact that in most instances the man's own word as to his previous occupation was taken. However, with all due allowance the general trend is



rather striking. It seems reasonable that the intelligence requirements of the professions should be more exacting than those of the unskilled laboring jobs and that the figures given indicate something like the intelligence requirements of the occupations in question.

The army data have subsequently been gone over carefully (190) for 3600 cases, making corrections for possible sources of error, such as exemptions, and giving for 96 occupations the average intelligence, and the range of the middle fifty per cent, just as is done in Table XXVIII. The occupations are grouped under the following classification:

1. *Professional* (superior intelligence required).
  - A. Very high standards.
  - B. Slightly lower standards — professional and educational.
2. *Technical* (high average intelligence required). Technical work, business, promotion, clerical, highly skilled mechanical work demanding leadership.
3. *Skilled mechanical work* (average intelligence required).
4. *Semi-skilled and low-skilled* (low average intelligence required).
5. *Unskilled* (inferior intelligence). Manual work, no skill.

The general trend of results is the same as that just discussed. For instance, averaging in class A intelligence (using army standards) are civil and mechanical engineers and clergymen; in class B are physicians, teachers, chemists, draftsmen, dentists, and minor executives; in C + are stenographers and bookkeepers; in C - at the upper end are some of the skilled workers such as masons or shoemakers, and going down toward the lower end of the class we find cooks, textile workers, and sheet-metal workers, with laborers at the bottom, while fishermen are the only ones that fall in class D intelligence.

A study of a group of men applying at a bureau for vocational advice is worth passing mention. Using the foregoing classification in its entire detail, the occupations were divided into ten different classes on the basis of the requisite intelligence. Each of the 300 men was put into one of these classes on the basis of his previous occupation or his occupation at the time of application. He was also given the army intelligence test. It was then noted

## INTELLIGENCE AND VOCATIONAL APTITUDE 269

whether his actual intelligence fell in the same class as that possessed by the average person in his occupation, or whether he was too high or too low for the occupation. It developed that only 14 per cent of the group fell in the same tenth in intelligence that they should have if they were ideally placed according to the above assumptions; 69 per cent had more intelligence than that required by their job, and 17 per cent had less. If we make a more liberal allowance and consider as properly placed those who are in the right class or in one class above or below it, these individuals constitute 31 per cent of the group. These results simply indicate that a great many workers are vocationally misplaced, and that present methods of hiring workers or determining vocational objectives have fallen short in many respects.

**Miscellaneous occupational groups.** Since the war various intelligence tests have been given to miscellaneous groups of workers. In instances where the same test has been given to a considerable number of groups, it is possible to note the average intelligence of

TABLE XXIX  
AVERAGE INTELLIGENCE SCORES OF OCCUPATIONAL GROUPS <sup>1</sup>

College presidents (small colleges).....	58
Engineering students.....	57
Students in medical college.....	56
Students in arts college.....	54
Sales executives.....	54
Supervisors in manufacturing plant.....	52
Executives of progressive firms.....	51
Rotary club members.....	46
Nurses.....	42
Foremen.....	41
Office employees.....	40
Machine operators.....	33
Sales force in department store (men).....	33
Office boys.....	31
Sales force assisting in holiday rush (men).....	29
Students in business college.....	28
Sales force in department store (women).....	27
Sales force assisting in holiday rush (women).....	25

<sup>1</sup> After Scott and Clothier.

each group in the same fashion as in the foregoing cases. (525.) The results of such a study are given in Table XXIX. The figures indicate the average score of the group under consideration in actual points in a test which has a maximum possible score of 113 points. The results are quite similar to those obtained with the army tests. The retail salespeople have the lowest intelligence of those studied. Machine operators, office employees, and foremen are somewhat superior. Rotary club members, who are presumably successful business men, are higher still. Executives and college students make still better scores, and college presidents are at the top. This hierarchy does not run down as far as the unskilled laborers, but the presumption is that their average score would be still lower than that of any of the groups listed.

Another intelligence test was given likewise to a considerable

TABLE XXX. INTELLIGENCE SCORES OF OCCUPATIONAL GROUPS <sup>1</sup>

	FIRST QUANTILE	AVERAGE	THIRD QUANTILE
Major executives.....	90	127	156
First-year graduates (business).....	109	125	140
Sales engineers.....	110	120	150
College seniors.....	100	118	137
School superintendents and special subject teachers.....	100	109	119
Executives (general).....	82	102	116
Real estate salesmen.....	80	102	115
Office specialty salesmen.....	60	95	112
Students in school for insurance salesmen.....	74	93	105
Experienced insurance salesmen.....	60	86	110
Office clerks.....	55	84	105
Semi-specialty salesmen.....	56	78	106
Routine salesmen.....	41	71	94
House-to-house salesmen.....	30	65	95
Trade high school (night).....	37	62	87
Policemen.....	*	42	*
Retail sales clerks (notions, bargain counters, etc.).....	20	33	50

\* Figures not available.

<sup>1</sup> From Kenacy and Yoskum's *The Selection and Training of Salesmen*, by permission of the McGraw-Hill Book Company, Inc., New York.

## INTELLIGENCE AND VOCATIONAL APTITUDE 271

number of occupational groups. (278, 276.) The results are presented in Table XXX. The first quartile, the average, and the third quartile are given as in Table XXVIII. A few of the groups involved are rather selected and hence score more highly than would similar groups taken at random. This is particularly true of the major executives, the first-year graduates in business school, and the real estate salesmen. The retail clerks likewise are confined to a group selling a particular class of commodities and a more random selection would probably make lower scores. The hierarchy, however, is rather obvious ranging from major executives, engineers, and students down through school teachers, general executives, special groups of salesmen, office workers, routine salesmen, and policemen to the retail clerks at the bottom of the intelligence scale.

A notion of the functioning of this hierarchy at the upper end of the intelligence range may be obtained from a consideration of the results of the army test given to the entire student body of a large university and a majority of the members of the faculty who taught those same students. The average for the students is 136 points actual score in the test and for the faculty 154. The average for the entire army was about 60 points. The middle half of the students

TABLE XXXI. PER CENT OF FACULTY AND STUDENTS IN DIFFERENT INTELLIGENCE CLASSES

CLASS	FACULTY	STUDENTS	ARMY
A	77	51	5
B	16	33	10
C+	4	13	18
C	2	2	29
C-	0.4	0.2	21

fall between scores of 116 and 155, while the middle half of the faculty scores fall between 139 and 174. The figures are expressed in terms of the army letter grades in Table XXXI. According to

those standards, A meant very superior intelligence, while C meant average. The table gives the per cent of faculty and of students in each of these classes and also for comparison, the per cent of soldiers in the draft falling in these same classes. This table shows the superiority of the faculty to the students, and in turn the superiority of both to the unselected men in the army, who probably represented the whole gamut of occupations.

The results should perhaps be somewhat qualified by the fact that the examination was voluntary with the faculty, but compulsory with the students. It has been found in other connections that persons willing to take an intelligence test, or at least willing to write their names on their test blanks, grade somewhat more highly than those reluctant to do so. This would tend to lower the faculty results somewhat if the entire group had been involved. However, with differences of the magnitude shown in the table, there is clear indication that those in the profession of college teaching stand higher in the intelligence hierarchy than their students, many of whom will ultimately settle in much less intellectual types of occupation.

In another instance a brief intelligence test was given to various occupational groups. (552.) The average scores and the range, i.e., maximum and minimum scores of the groups,<sup>1</sup> are given in Table XXXII. A similar tendency is manifest, for occupational

TABLE XXXII. INTELLIGENCE SCORES OF OCCUPATIONAL GROUPS <sup>2</sup>

	LOWEST	AVERAGE	HIGHEST
Prospective executives (students) . . . .	54	63	70
Retail shoe salesmen . . . . .	30	40	50
Buyers for retail store . . . . .	22	38	53
Waitresses . . . . .	12	22	30
Retail saleswomen . . . . .	10	16	22

groups, to differ markedly in intelligence. The retail saleswomen and the waitresses are distinctly inferior to the buyers and retail

<sup>1</sup> The figures are only approximate, as they were estimated from a bar diagram published in the original account of this work.

<sup>2</sup> After Starch.

## INTELLIGENCE AND VOCATIONAL APTITUDE 273

salesmen, while the prospective executives are far superior to any of the other classes involved.

**Different departments in an organization.** A set of tests, the total of which was tantamount to an intelligence test, was given in a rubber-tire plant. The average scores of various occupational groups within this one concern appear in Table XXXIII. The

TABLE XXXIII. AVERAGE INTELLIGENCE SCORES OF OCCUPATIONAL GROUPS IN ONE CONCERN

Laboratory and drafting.....	147
Factory council.....	144
General clerical workers.....	138
Shipping department.....	112
Factory committee.....	108
Foremen.....	88
Inspectors.....	86
Finishers and builders.....	87
Handing out stock.....	76
Truckers and mixers.....	47

highest scores are attained on the average by a group of employees in the laboratory and drafting departments. These individuals are, of course, technically trained. Slightly inferior to these, but perhaps not significantly so, are the members of the factory council, a group of six executives who, at that time, determined the policies of the organization. Below these come a group of general clerical workers, who compare rather favorably with the executives, and are distinctly superior to the other groups involved. The next in order are the employees in the shipping department, followed closely by members of the general factory committee. This committee was comprised of a few foremen and various minor executives who met regularly to determine less important questions of policy. Between this group and the foremen and inspectors there is a considerable gap. The men who finish and build tires compare favorably with the foremen and inspectors. This probably reflects the well-known fact that foremen are in some concerns chosen, not by virtue of any superior capacity, but simply because they are experienced workmen. The employees who hand out

stock are somewhat inferior to the finishers and builders and foremen. Far down at the bottom of the scale are those employees engaged in unskilled labor, such as hauling trucks or mixing and washing crude rubber.

**Different types of salesmen.** The foregoing discussion has dealt with the occupational hierarchy for the whole range of occupations from unskilled labor to the professions. The question arises whether there is any such hierarchy within a given occupation. Some data for salesmen are available on this point. A number of the occupational groups listed in Table XXX (*supra*) may be classed as salesmen and there is some evidence of a hierarchy. The sales engineers make the highest scores in intelligence. The real estate salesmen are appreciably lower. A little lower still are the office specialty salesmen and the students in a school for insurance salesmanship. The experienced salesmen are inferior to the students, but the two groups manifestly overlap very considerably. Next in order come the semi-specialty salesmen, then the routine salesmen, with the house-to-house salesmen lower still. At the bottom of the scale, far inferior to any of the others in intelligence, are the retail sales clerks.

Quite similar results were obtained in another study of four different groups of salesmen. (384.) The results are presented in Table XXXIV. Its form is identical with that of Table XXX, and

TABLE XXXIV. INTELLIGENCE SCORES OF GROUPS OF SALESMEN<sup>1</sup>

	FIRST QUARTILE	AVERAGE	THIRD QUARTILE
Salesmen for technical product. . . . .	124	139	155
Insurance salesmen. . . . .	82	112	138
Wholesale salesmen. . . . .	59	89	121
Counter salespeople. . . . .	36	51	70

<sup>1</sup> After Miner.

it gives the average intelligence of each group as well as the first and third quartile score. The same tendency is manifest. The

men who sell a highly specialized technical product stand at the top, while the counter salespeople are at the bottom. There is considerable overlapping, especially of the wholesale and insurance groups, but there is sufficient difference to be of interest.

It would appear that even within a single vocation, such as selling, there is an intelligence hierarchy. All salespeople have considerable in common in that they are inducing prospects to purchase something. But it appears that even with this common element certain types of selling actually require a higher order of intelligence than do others. These results should not be interpreted to mean that intelligence tests alone will give a good prediction of selling ability. Neither do they imply anything about the diagnostic value of intelligence tests within a particular group of salespeople, such as retail clerks. They do indicate, however, that over and above the other mental qualifications requisite for salesmanship certain aspects of this vocation are more exacting in their intelligence requirements than are others.

The theory underlying the various results just presented is that a person will in the long run tend to get about as high in the occupational scale as his intelligence warrants. If he attempts a job too high in the scale, he will find it too exacting and leave either voluntarily or involuntarily. If, on the other hand, he starts with one that is too low in the scale, he will not find it sufficiently interesting because it does not afford an adequate outlet for his intellectual ability, and he will leave it for something higher. The result is that he ultimately lands at about the maximum level at which he can do effective work. Other factors will, of course, sometimes alter the results. A lazy person may not want a more exacting type of work and a person with unattractive appearance or some personality defect may be refused a job for which he is capable. The above assumption deals only with the average case. If it is valid, we may then conclude that the average intelligence of an occupational group indicates approximately the degree of intelligence that is necessary for effective work in that occupation.

These principles may then be used to some extent in the practical problem of employment. If we know, for instance, that persons below a certain intelligence score, such as C + on the army standard,



are seldom found in clerical or executive positions, it will probably be well, in lieu of further special examinations, to select for such positions persons with intelligence at least equal to C+. The lines cannot be drawn too closely, but extreme values surely are significant. Persons of very low intelligence, such as that possessed by the average unskilled laborer, would doubtless be distinctly misplaced when put in an executive or clerical position, and it would presumably be better policy to give them unskilled or semi-skilled laboring work. By this procedure we cannot hope to predict an individual's success in a given line of work in terms of probability as is possible when a correlation coefficient is available. The most that we can do is to locate the individual at somewhere near his occupational level. This information, however, is often valuable and is especially so when dealing with extreme cases of discrepancy between the intelligence possessed by the applicant and that required for a given occupation.

#### INTELLIGENCE SURVEYS

It is sometimes profitable to conduct a survey with intelligence tests throughout an organization, or a group of similar organizations. This may often reveal conditions that were unsuspected. Quite apart from devising methods of predicting occupational success, it is often of interest to determine what has happened up to date with the usual employment methods. If the survey is conducted on a rather large scale, sampling a very considerable range of jobs within the concern, it is quite probable that the usual hierarchy will be found as was the case with the concern surveyed in Table XXXIII. Other factors are, however, sometimes brought out in such surveys and a few typical results are given below. One cannot tell in advance just what to expect, but often something will turn up that will throw rather interesting light on employment problems. (307.)

**Male vs. female employees.** A company had a large number of office employees of both sexes. It occurred to some one to compare the intelligence of the two groups. In the particular test used, the male office employees averaged 51 points and the female employees averaged 38 points. This casts no reflection on the intelligence of

women in general. It merely shows that the company had selected for its office a somewhat higher grade of men than of women. It is possible, too, that some men of high intellectual capacity take a clerical position as a stepping-stone to executive work. At any rate, the results indicate the desirability of judging male office employees by standards derived from testing men and *vice versa* in judging female employees.

**Similar employees in different companies.** A survey was made of the women office employees in several different companies. In one company their average intelligence was 31 points, in another 38 points, in another 42, and in a fourth 46 points. Obviously the companies had different standards and some were more exacting than others. A similar situation was found with reference to office boys. In one company their average intelligence was 26, in another 32, and in a third 36 points. Evidently the last company was employing a higher type of personnel for this work. The boys in this last concern would manifestly form a better source from which to recruit future executive material.

**Applicants *vs.* employees.** In two concerns where the intelligence of the office employees had been determined, similar tests were given to all the applicants for office work. In the first concern the applicants averaged 36 points in intelligence and the employees 29 points, while in the second concern the applicants averaged 38 and the employees 47. Evidently both concerns were attracting about the same kind of applicants. The second concern, however, was employing a much higher type of personnel. To analyze this difference it would be necessary to know more about the employment methods of the companies and their wage policies. There were preliminary indications that the first company selected high-grade individuals, but failed to keep them because they left for more lucrative positions elsewhere. Data of this sort raise at least the problem of further analysis of the policies and methods of the companies.

#### CORRELATION OF INTELLIGENCE WITH THE CRITERION

The foregoing methods are not the only ones by which the problems of intelligence and employment may be approached. The technique described in previous chapters for evaluating special

capacity tests is applicable likewise with reference to intelligence tests. A group of persons engaged in a certain occupation may be given intelligence tests and then their test scores correlated or otherwise compared with the criterion. The statistical technique is exactly the same as described previously so that no further discussion of it will be given in the present connection. In some of the cases to be described, the criterion consisted of production figures or careful estimates made by the employees' superiors and correlation coefficients were computed. In other cases a less refined comparison was made of different groups of workers.

**Clerical workers.** With a group of women office workers the correlation of intelligence and the supervisor's estimate as to the worker's ability was .76. (307.) With a similar group in another company the correlation was .82. These are comparatively high correlations. In one other group the correlation proved to be only .40. Subsequent analysis revealed, however, that the supervisor had rated the women on the basis of length of service rather than on actual proficiency. It would seem that intelligence is one of the main requisites for this kind of work or at least that those who have high intelligence possess the other necessary qualifications.

**Office boys.** With a group of messenger boys, those discharged averaged 22 points in an intelligence test, while those promoted averaged 39 points. In another case a group was tested and the results filed for twenty-one months. At that time the average score of those who were still in the company was 42 and of those who were not in the company 35. Further analysis of those who were no longer with the company revealed that those who left to accept better positions averaged 45 points and those who were discharged averaged 28 points. Of those who remained the ten boys who stood highest in the test were receiving an average salary of \$16, while the ten who stood lowest in the test were receiving an average of \$13.40. (525, 266.) The executives under whom these boys worked estimated their future value to the company by classing them into four groups as follows:

- A. Probable high-grade executive ability.
- B. Probable minor executive ability.
- C. Without executive ability, but good clerical timber.
- D. Probably best adapted for highly mechanical job.

## INTELLIGENCE AND VOCATIONAL APTITUDE 279

The average intelligence and the average salary of each of these four classes is given in Table XXXV. It is to be noted that the

TABLE XXXV. INTELLIGENCE OF OFFICE BOYS <sup>1</sup>

EXECUTIVE'S ESTIMATE	AVERAGE SALARY	AVERAGE TEST SCORE
A	\$16.78	66
B	\$14.48	55
C	\$13.74	51
D	\$14.26	39

<sup>1</sup> After Scott and Clothier.

executive's ratings and the test scores agree perfectly. There is likewise a fair agreement between salary and the other two factors.

**Clothing operators.** The production of operators in clothcraft shops was correlated with intelligence. The coefficient was .51. In using the test subsequently some persons with low scores were hired, but were assigned to less exacting work. The conclusion was drawn that "in clothcraft shops the use of mental tests, although only a partial measurement, is the quickest, most accurate, and most economical method of prophesying future skill at machines and of placing operators at types of work suited to their capacity." (525, 266.)

**Executives.** An intelligence test was given to minor executives in 1915 and then in 1920, the results compared with their firm rank. The correlation was .69. A small group of executives at the head of a concern were ranked by the vice-president as to their executive ability. The correlation with their rank in an intelligence test was .89.

When we consider business success in general rather than executive ability within a single organization, a somewhat different result is obtained. A group of business men at a conference took an intelligence test. (64.) They were subsequently sent a questionnaire dealing with their business career, and on the basis of these questionnaires five judges rated them as to "success." The judges

agreed fairly well among themselves as shown by an average correlation of .60 between the different judges. The combined "success" rating correlated with intelligence to the extent of  $-.10$ . The conclusion is drawn that "the evidence in hand suggests that superiority in intelligence above a certain minimum contributes relatively less to business success than does superiority in several non-intellectual traits of personality."

**Salesmen.** While the foregoing results have indicated in most cases some correspondence between intelligence and occupational efficiency, it is unsafe to generalize and conclude that this is true of all occupations. Many instances are found in which the results are not so clear-cut. Salesmanship is one of these. The results are somewhat equivocal, but in general the relation of intelligence to selling ability is slight. With two groups of retail sales clerks the correlations between managers' ratings and intelligence were  $-.11$  and  $-.26$ . (278, 260.) This indicates a small inverse relation between intelligence and the criterion. In fact those who were rated the highest in efficiency were appreciably below the average in intelligence. On the other hand, a group of shoe salesmen were classed by executives as good and mediocre. (552.) The former ranged from 33 to 59 in test score and the latter from 19 to 44. Similarly the saleswomen in the same establishment were rated as above-average, average, and below-average. The average scores of the three groups were respectively 95, 71, and 41, although there was more overlapping of the groups than in the case of the men.

For house-to-house salesmen there was found a zero correlation between production and intelligence. It seemed that a man with low intelligence stood as good a chance of success in this line as did a man with high intelligence. (278, 261.) With two groups of routine salesmen the correlations were respectively  $-.06$  and  $.00$ . There was, however, a little indication that those of lower intelligence were better than those of high intelligence. For the men who were above average in production the average score was 64, for those who were average in production the score was 65, and for those below average in production the score was 78. Similar results were found with heating-equipment salesmen. The correlation was insignificant, but the average scores for above-average, average, and

below-average salesmen were respectively 74, 72, and 94. This may have been due, however, to the fact that a considerable number of high-grade men had been recently employed and had not had sufficient time to demonstrate their ability.

Of a large group of life insurance men the sales managers averaged 93 points in a test and the whole group of salesmen 83 points. Promotion to managership in this field usually depends on success in selling, so there was some indication of the value of intelligence. In a smaller group the correlation of intelligence with two-year production was .24 and in another group the correlation with four-year production was .34. In a single company the correlation of intelligence and production for a small group was .60.

In two companies the office specialty salesmen showed a very slight correlation, but there were some indications of relationship when the managers were considered in comparison with the salesmen. The average intelligence scores of managers, active salesmen, and inactive salesmen in the first company were 76, 73, and 69 respectively, and in the second company 74, 69, and 73. In so far as promotion to the position of manager indicates success, there is a slight indication of a positive relation between intelligence and success in selling this specialty.

These results do not conflict with those presented earlier regarding the intelligence hierarchy. It was shown there that certain types of selling are somewhat more exacting from the standpoint of intelligence than are others. But when we consider salesmen of a given sort the results are not very clear-cut. There is some indication that in the lower grades of selling, such as retail clerking, there is a slight negative relation between intelligence and proficiency, while at the upper end, such as insurance or specialty selling, there is a slight positive relation. The small amount of these relations may be in part due to the fact that salesmanship appears to be in a period of transition from selling through individual efforts to selling through advertising, so that the work of the salesman is at present less definite and measurable. Production figures for selling, moreover, are influenced by extraneous factors, such as territory, to a greater extent than are similar figures for workers in a factory. At any rate, it is more difficult to predict selling ability on the

basis of intelligence than it is to predict some of the other occupational abilities above mentioned.

**Silk mill operatives.** A large number of employees in a silk mill were given various intelligence tests mostly of the performance type. The correlation between tests and production was practically zero. "The best weaver in the mill took 10 minutes to assemble a puzzle that an intelligent person does in 25 seconds." (435.) It seemed that in the work where the machinery was automatic and little skill needed, high intelligence was not required and might even be detrimental. It is quite possible that a person of high intelligence will revolt at such monotonous work, and that one requires rather stolidity, patience, inertia of attention, regularity of habits, and other temperamental rather than intellectual traits.

**Operations in industrial school.** The boys in various occupational groups at an industrial school were rated in proficiency relative to the others in that same trade. (137.) They were given Binet tests and mental age was correlated with trade rating. In most instances the correlation of intelligence with the criterion was small. However, there were a few cases of appreciable positive correlation coefficients, and also a few negative coefficients. Some of these are given in Table XXXVI.

TABLE XXXVI. CORRELATIONS OF INTELLIGENCE AND TRADE ABILITY IN INDUSTRIAL SCHOOL <sup>1</sup>

Office . . . . .	.98
Poultry . . . . .	.60
Hospital . . . . .	.41
Printing . . . . .	.33
Gardening . . . . .	-.23
Laundry . . . . .	-.30
Bookbinding . . . . .	-.31
Shoe shop . . . . .	-.31
Plumbing . . . . .	-.38

<sup>1</sup> After Cowdery.

Office work shows a very large correlation. This suggests similar fairly large coefficients mentioned above for clerical workers. The poultry department likewise shows a fairly high coefficient fol-

lowed by hospital and printing work. On the other hand, a few negative coefficients are manifest. The largest of these is for plumbing and the next in order for shoe-shop, bookbinding, and laundry work — all on par. While these coefficients are not large, their existence is suggestive. It is possible that in some of these types of work greater proficiency goes with lower intelligence — provided proper supervision is given. It is probable that the boys in an industrial school are supervised more carefully than the average adult in industry. Consequently the correlations might not be so large in the ordinary practical situation.

The obvious implication of these studies of vocational proficiency as compared with intelligence is that intelligence tests are valuable in selecting employees for some kinds of work, but that for other kinds of work they are worthless. It is an unwarranted assumption that for a particular job the most intelligent person available is to be preferred. Just as in dealing with tests for special capacity it is necessary to test the tests, so, in dealing with intelligence as predictive of ability for a given occupation, it is necessary first to correlate or compare in some way efficiency in the test with efficiency in the job. As far as intelligence tests have been employed in industry, they have proved most useful (aside from locating workers at approximately their appropriate level in the hierarchy of occupations) in selecting clerical workers, office boys, and executives.

#### CRITICAL SCORES IN INTELLIGENCE

**Method.** If it is established that intelligence is related to proficiency in a certain job and the tests are to be used for employment purposes, the problem arises of establishing a critical score as a basis for hiring or rejecting applicants. The procedure here is identical with that used in the case of tests of special capacity. The most probable ability in the job may be computed from a regression equation or by the use of distributions like Table XVII, and then a decision made as to how big a chance it is desired to take. Or the critical score may be set by inspection of the data — comparing extreme cases — or by determining in a scatter plot where a line can be drawn with the least overlapping of two classes of vocational ability. The method will not be repeated in detail, but



reference made merely to the description in the previous chapter. A few examples of critical scores determined in one or another of the usual methods will be cited by way of illustration.

**Examples.** In a large tire manufacturing concern in which intelligence tests were rather extensively used, critical scores were established for a considerable number of jobs. (525, 242.) Some of these are given in Table XXXVII. This table suggests some of

TABLE XXXVII. CRITICAL INTELLIGENCE SCORES IN A TIRE CONCERN <sup>1</sup>

<b>Women:</b>	
Stenographers.....	35
Typists.....	33
Comptometer operators.....	27
Clerks.....	23
<b>Men:</b>	
Factory school instructors.....	50
Chemical engineers.....	45
Other engineers.....	40
Draftsmen.....	35
Clerks.....	30
Dispatch clerks.....	30
Inspectors and foremen.....	23
Messenger and mail boys.....	15

<sup>1</sup> After Scott and Clothier.

the earlier ones presented in discussing the occupational hierarchy. In that connection it was the average intelligence of each occupational group that was of interest. In the present connection it is rather a minimum intelligence, below which the person has little promise of success. Usually a person below the critical score is not hired unless he has some compensating qualifications.

In the study of office boys mentioned above, it was decided to set a critical score of 32 points. On this basis only 43 per cent of those below this score remain with the company, while 62 per cent of those above this score remain. The group below 32 points contains only 1 of the 29 boys who were promoted and all 16 of those discharged.

In the study of shoe salesmen above mentioned, a critical score

of 33 points would rule out none of the good group and would eliminate 57 per cent of the mediocre group.

In office specialty selling a score of 50 seemed to be critical. All the managers scored above this figure. Consequently, in employing prospective managerial material persons above this score were selected. In one company, of the 19 men below this critical score 7 left the employ, 8 produced very little, and 2 of the remainder were below average in production. (278, 265.)

In connection with work of a vocational adjustment bureau (98), critical scores for a number of types of work were determined on the basis of mental age. For instance, in millinery work girls whose mental age was 9 and 10 years seemed adapted to such work as sewing linings in hats, or steaming material. A mental age of 11 was necessary for an improver's job, i.e., an operation in which the foundation of the hat is covered and a wire edge attached. A mental age of 12 appeared necessary for machine work on straw or other material. In another instance a critical score of about  $7\frac{1}{2}$  years qualified one for packing powder puffs, whereas for packing articles which needed to be separated or folded, such as hair-nets, a mental age of  $9\frac{1}{2}$  was established as a minimum. Similarly in hand sewing, a mental age of 9 was sufficient for mounting buttons on cardboard, whereas hand sewing garments necessitated an age of 10 and sewing labels an age of 12. A mental age of  $9\frac{1}{2}$  years sufficed for cut-out or pasting work,  $10\frac{1}{2}$  for keeping stock or checking, whereas 13 was required for assembling more complicated parts. These critical scores were not established as hard-and-fast lines, but were useful in making rough vocational adjustments.

#### OPTIMUM vs. MAXIMUM INTELLIGENCE

It might be supposed that in a given vocation which showed some relation between intelligence and success, it would be advisable to hire persons with the maximum intelligence. A critical score might be established for the minimum intelligence that would enable one to do satisfactory work, but above this critical score it might be supposed that the more intelligence possessed by the applicant the better. Recent work, however, has shown that in some cases there is an upper critical score as well as a lower. In other

## EMPLOYMENT PSYCHOLOGY

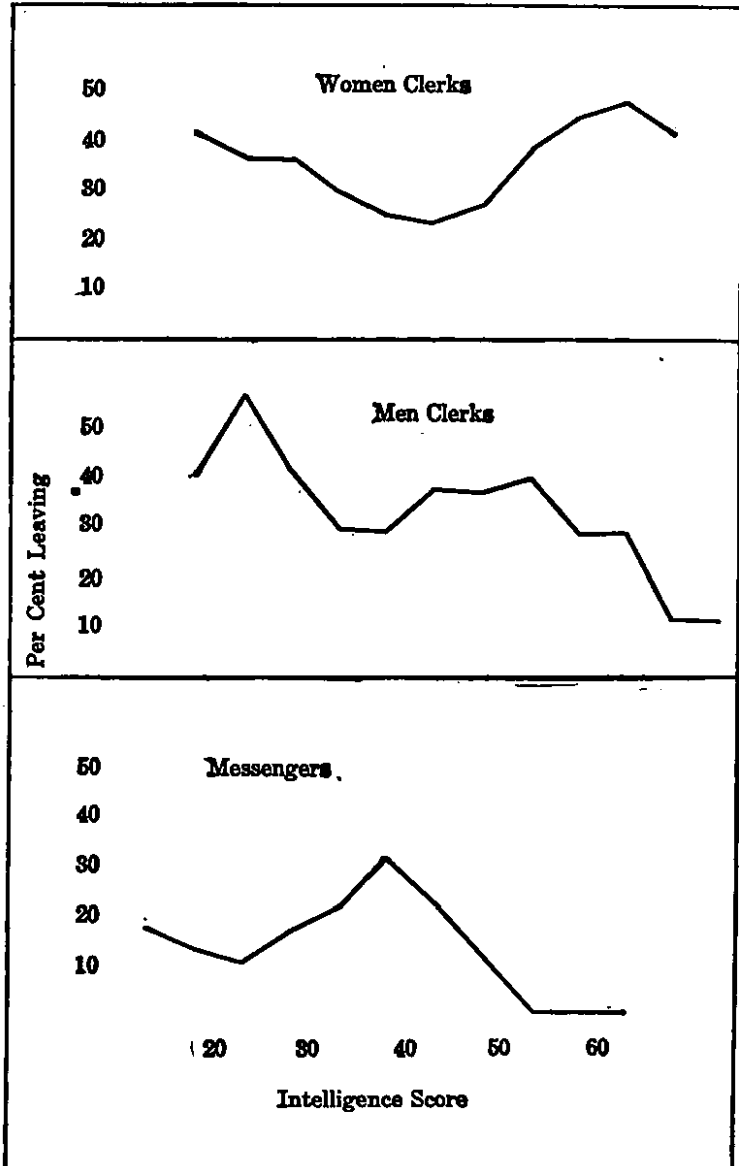


FIG. 7. INTELLIGENCE AND OCCUPATIONAL STABILITY

words, what we need is not prospective workers with *maximum* intelligence but rather with *optimum* intelligence. These facts come out clearly in studies of turnover or permanency in relation to intelligence and reveal that a person may be too intelligent for his job so that it fails to interest him and he consequently quits.

**Stability and intelligence: office workers.** In a survey of an office force, stability was plotted against intelligence. (307.) These results are shown in the two upper curves of Figure 7. Along the base line are the test scores. The vertical distances represent the per cent of those with the given score leaving the job within six months from the time they were hired. The results are most striking in the case of the women clerks. Those with scores between 30 and 50 are more stable than the others. A large per cent of those with low intelligence leave, presumably because they do not have sufficient ability to be effective in this line of work. But there is likewise a large per cent of those with high intelligence who leave. It is probable that the job is not sufficiently exacting to hold their interest. High intellectual capacity apparently demands expression or exercise, and they are discontented. Of course other factors may play some part, but contentment is probably no mean factor.

Similar results were found in another company. Between 40 and 50 per cent of the women clerks in the office with high or low intelligence left within six months, whereas about only half as many with medium intelligence left in that period. In another case there was a correlation of  $-.45$  between intelligence and length of service. This means that the more intelligent worker left earlier than the less intelligent.

In a large clerical force turnover was computed for a period of 30 months. (45.) The work was graded into five degrees of difficulty denoted by A, B, C, D, E, — A being the lowest grade of clerical work. Two arbitrary points in the intelligence scale were selected — 80 points and 110 points, and for each grade of work the turnover computed for those below 80 and over 100. The results are shown in Table XXXVIII. One notes immediately that for low-grade jobs (A and B) the most intelligent workers have the highest turnover; while for the lowest-grade job (A) the least intelligent are the most

TABLE XXXVIII. TURNOVER FOR CLERICAL WORKERS OF HIGH AND LOW INTELLIGENCE <sup>1</sup>

GRADE OF WORK	PER CENT TURNOVER FOR INTELLIGENCE LESS THAN 80 POINTS	PER CENT TURNOVER FOR INTELLIGENCE OVER 110 POINTS
A.....	37	100
B.....	62	100
C.....	50	72
D.....	58	53
E.....	66	41

<sup>1</sup> After Bills.

stable. In still another company 40 per cent of the clerks scoring less than 30 points in the test left within six months. This per cent decreased up to about 50 points in the test, then increased again for those who made higher scores. (542.)

**Messenger boys.** Results at variance with the foregoing were obtained with a group of messenger boys. (525, 253.) The results are shown in the lowest curve of Figure 7. The fewest resignations occurred among the boys with high and low scores. For the low group this was perhaps due to the fact that the applicants were sufficiently alert to hold the job, but incapable of improving themselves by going elsewhere. The data do not include boys who were discharged. The results for the high group were explained in this particular case by the fact that the work was not distasteful to the brighter boys, because it afforded them an opportunity to learn a good deal about the business and might serve as a stepping-stone to a higher position in the office. Many prominent executives have, of course, come up through the route of the office boy.

**Cashiers.** A group of cashiers and inspector wrappers were tested and results compared with stability. (636.) The facts are shown in Table XXXIX. It is obvious that the greatest stability is found in the middle range of intelligence.

**Policemen.** The army alpha test was given to a group of policemen in a large city. (604.) The average scores for different

TABLE XXXIX. INTELLIGENCE AND LENGTH OF SERVICE <sup>1</sup>

TEST SCORE	AVERAGE LENGTH OF SERVICE IN DAYS
10 to 19	3
20 to 29	91
30 to 39	156
40 to 49	142
50 to 59	107
60 to 69	100
70 to 79	96
80 to 89	87
90 to	35

<sup>1</sup> After Viteles.

groups are given in Table XL. The results shown in the first three lines of the table are not what one would ordinarily expect. One would suppose that the officers would have higher intelligence than the men under them. This is not the case. The remainder of the table, however, clarifies the matter. The more intelligent

TABLE XL. AVERAGE INTELLIGENCE OF POLICEMEN <sup>1</sup>

Lieutenants.....	58
Sergeants.....	55
Patrolmen (all).....	71
Patrolmen in service less than 9 months.....	72
Patrolmen in service 10 to 19 months.....	64
Patrolmen in service over 20 months.....	51

<sup>1</sup> After Thurstone.

patrolmen leave the service rather early. It is quite possible that the more intelligent patrolmen would have made better officers, but they did not remain long enough to get promoted. In another city the same tendency was found for the more intelligent to leave earlier, although the officers in this case made somewhat higher grades on the average than did the patrolmen.

Waitresses. A group of waitresses who had served 4 months to 15 years averaged 17 points in an intelligence test and their scores ranged from 4 to 33. At the same concern the waitresses who had served less than 4 months averaged 32 points in intelligence and

ranged from 15 to 45. Those of lower intelligence were manifestly more stable. (552.)

**Salesmen.** With retail clerks the correlation of intelligence and length of service with one group of employees was  $-.31$  and with another group  $-.11$ . (278, 266.) This gives a slight indication that the less intelligent ones tend to remain longer in the employ. With house-to-house salesmen no correlation was found. For routine salesmen a coefficient of  $-.44$  was obtained in one company and  $-.46$  in another. In the first of these groups only 30 per cent of those scoring over 70 points remained with the company  $2\frac{1}{2}$  years, while 64 per cent of those below 70 points remained for at least that length of time. Apparently the routine nature of the work, its easy mastery, and its lack of an attractive future produced instability among the more intellectual men. Heating-equipment salesmen showed similarly a correlation of  $-.26$ . Life-insurance salesmen, on the other hand, gave a small positive correlation (.23) between intelligence and length of service. The same thing was found with office-specialty salesmen. In one company the correlation for the sales managers was .61, and for experienced salesmen .21, while in another it was .12 for the managers and .50 for the salesmen. For the inactive salesmen there was, however, in one company a negative correlation of  $-.42$  between stability and intelligence. In general with the lower grades of selling there is a slight inverse relation between intelligence and stability, while with the higher grades there is a slight positive relation.

**Dissatisfaction and intelligence.** A bit of additional evidence as to the undesirability of too much intelligence in certain lines of work is obtained from a consideration of the attitude of different groups of workers and their varying degrees of satisfaction with their work. In a concern where considerable dissatisfaction was noted, it was analyzed with reference to the status of the most dissatisfied employees. Tests results were not available, but school retardation as manifested by age and grade at leaving school was noted as an indirect indication of intelligence. In the tool department where the work was fairly complex, the most dissatisfaction was noted among the workers who were presumably the most retarded intellectually. In the inspection department, on the other

hand, where the work was repetitive and monotonous, the retarded individuals showed the least dissatisfaction. The brighter individuals were apparently happier in the more complex work and the duller individuals in the simpler work. (542.)

In a school for unemployed young persons a number of women were stitching on a wide-meshed canvas of standard size and shape. The most intelligent ones experienced the most boredom and their output was the most variable. They could reach a high output, but would not maintain it. A girl of medium intelligence was the most effective worker and liked the work. A girl of low intelligence improved enormously in her work, but was disturbed by conversation. (96.)

**Upper critical score.** Such considerations as these have in some instances led to the use of an upper as well as a lower critical score. With the group of routine salesmen previously mentioned a critical score of 70 was established. (278, 262.) Scores *above* this were considered unfavorable. This was one of the cases of a negative relation between intelligence and selling. Only 37 per cent of the above-average salesmen scored over 70 points, while 62 per cent of the below-average salesmen exceeded this intelligence score. On the other hand, 63 per cent of the above-average salesmen scored less than 70, while only 37 per cent of the below-average salesmen fell below this critical score. In various other instances where the curve for stability takes the shape of the upper one in Figure 7, it has proved advisable to set a critical score at each end of the intelligence scale.

A person may be too bright for a given job just as he may be too dull. Such a person quickly masters the job, reaches its limits, and becomes dissatisfied. His work may be very effective almost from the start, but he "burns out" and leaves the organization. The desire for the most intelligent employee is sometimes due to the fact that inadequate training or supervision is given. An employee of high intelligence may be able to shift for himself more effectively at the outset, but he may not be so permanent an asset as the person a little lower in the intellectual scale. Vocational placement, then, does not involve merely the selection of the ablest man for the job as far as intelligence is concerned. Overstocking a low-grade job



with high-grade personnel will tend to increase turnover. In evaluating intelligence with reference to vocational aptitude, we should consider not merely maximum intelligence but rather optimum intelligence.

#### SUMMARY

There are occupations which do not require for their effective performance any specialized capacity, but rather general ability or intelligence. A consideration of the average intelligence of various groups of workers reveals an occupational hierarchy. The unskilled laborers are inferior in intelligence to the semi-skilled or skilled workers. These in turn are surpassed by persons in technical, business, or clerical work. Members of the professions come at the top of the scale. The theory is that a person will in the long run attain about as high an occupational level in the hierarchy as his intelligence warrants. Hence these group averages are tantamount to the intellectual requirements of the occupations in question. It is thus possible to locate an individual applicant at somewhere near the occupational level for which he is best fitted, and with applicants of extreme intelligence the assignment to occupations at the opposite extreme is manifestly inadvisable. Similar hierarchies are found for the various jobs within a single organization and for different types of salespeople. Retail clerks have the lowest average intelligence scores. They are surpassed by the wholesale and routine salesmen. These in turn are exceeded by real estate, insurance, and specialty groups, with salesmen for technical products and sales engineers at the top of the hierarchy. Apart from the other requisites of salesmanship, certain aspects of the occupation are more exacting in their intelligence requirements than are others.

Intelligence tests are sometimes useful in surveying an organization or group of organizations. Such a survey throws light on the results attained by present employment methods and often raises further problems for analysis. For instance, one company found that the male office employees possessed higher average intelligence than the female. Different concerns in the same community were employing clerical workers of distinctly different intelligence levels.

Several similar companies were attracting the same grade of applicants, but had marked differences in the resulting personnel. These findings pointed to the need for analysis of employment methods and policies.

Intelligence tests have been in many instances compared or correlated with proficiency in various occupational lines. Fairly high correlations have been found with clerical workers, office boys, operators in clothcraft shops, and with certain types of executives. The results with salesmen were more equivocal. There were indications of small negative correlations with intelligence for the lower grades of selling ability and small positive correlations for the higher grades. In some other occupations no correlation whatever has been found and in a few instances of rather closely supervised work appreciable negative correlations.

Critical scores for intelligence may be set in the same fashion as critical scores for special ability tests. Some concerns maintain a set of critical scores for different jobs in their organization, especially office jobs. Workers falling below these critical points are not hired unless they possess some compensating qualifications.

In occupations which show a correlation between proficiency and intelligence, it is not necessarily desirable to employ persons with the maximum possible intelligence. Such individuals may learn readily and become effective workers soon after their induction, but in many instances it has been demonstrated that they do not remain long in the employ. With various types of office workers, cashiers, policemen, waitresses, and some of the lower grades of salesmanship, there has been found to be more instability or turnover among those of high intelligence than among those of average intelligence. While persons of very low intelligence may not have sufficient ability to learn effectively and perform their duties, those of very high intelligence may be too good for the job. It is not sufficiently exacting to hold their interest, their intellectual ability has insufficient outlet, and they become dissatisfied. This points in some instances to the necessity for an upper critical score. Applicants scoring above this amount are considered unsuitable material from the standpoint of permanency. Where intelligence is related to vocational aptitude, it is often desirable to consider not maximum intelligence but optimum intelligence.

## CHAPTER XI

### INTERESTS IN EMPLOYMENT PSYCHOLOGY

OCCUPATIONAL success depends on many things in addition to innate ability. Any employment man will immediately recall instances in which an applicant with the requisite ability was an occupational failure because he did not use that ability. Such persons are the bane of the psychologist's existence. In the initial stages of his work they raise havoc with his correlations between test and job. After his methods have been developed, he will frequently predict an applicant's success on the basis of test score and then the man will fail to come up to expectations. This failure to exert himself may have been due to the man's lack of incentive or lack of interest. The former of these lies without the field of employment psychology. The problem does not arise until after the man is hired and it involves the consideration of methods of instruction, working conditions, wages, and various other incentives which motivate the worker. The problem of interest, however, is germane to the present discussion. Many a man is physically present in his work, but mentally absent. This, of course, is undesirable, for he is less apt to use his capacities effectively and is more apt to be discontented. To be sure, it is sometimes possible to modify an interest or to arouse one where it has not existed previously, but many applicants approach the prospective employer with pretty firmly established interests. Whether these are innate or acquired is of minor importance compared with their firmly fixed character. They give the worker a certain bias which may or may not be favorable to his success. Consequently the study of interests is a logical aspect of the employment program.

It is rather obvious that wide differences in interests exist between individuals. A casual consideration of one's acquaintances will reveal this. Some persons enjoy tinkering with tools or machinery, while others dislike to drive a nail. Some enjoy meeting people and talking with them, while others are content with

their own company. Some enjoy classical music, while others prefer jazz. Some are enthusiastic about art or literature, while others give it little attention. Some are scientifically curious about the reason for the things in their environment, while others are content to take things unquestioningly as they find them. Some of these individual differences in interest may be of vocational significance. It only remains to devise more effective means of ascertaining their existence and of evaluating their practical importance once they have been discovered.

#### PERMANENCE OF INTEREST

Many workers come to the employment office with interests that are apparently rather firmly fixed. This question of permanence of interest has been studied statistically. (594.) About 350 individuals were requested to estimate in retrospect their relative interests in certain school subjects—mathematics, history, literature, science, music, drawing, and manual work. They estimated their relative interest in these subjects in grade school, then in high school, and finally in college. While errors of recollection doubtless enter into such estimates, the results were sufficiently striking to carry a presumption of some permanence of the interests. Correlations from .60 to .70 were found between interest at the age of 10 to 14 and at the age of 21. These interests, of course, involved only academic subjects, but some of them—for instance, the manual interest—might be of vocational significance. Furthermore, it is probable that similar permanence of interest would be found if other types were investigated in the same way.

A group of college women before graduation expressed their vocational interests, indicating in a list of vocations five choices in order of preference. Two years later they were sent a questionnaire asking for a similar record of vocational preference. Many of them had meanwhile had opportunity for their subsequent employment to alter their initial interest. However, seventy-five per cent of them still maintained the same vocation as their first choice, although forty-one per cent changed their second choice.

A study of high-school seniors, on the other hand, indicated that about half of them had changed their vocational intention at least

once prior to that time. (146.) Even in a supposedly stable group, such as appears in *Who's Who*, sixteen per cent changed their vocations at some time, due, presumably, in many cases to a shift of interest.

There is thus sufficient indication of permanence of interest to make it worth the consideration of the employment psychologists. It seems at least characteristic of the majority of individuals, although in many cases there may be a shift. The layman is perhaps inclined to overestimate this permanence. A parent is going entirely too far in assuming that because the child plays with a toy train his destiny lies in a locomotive cab, or that his predilection for filling bottles with water presages an adult interest in pharmacy. Almost every boy at some time looks forward to becoming a policeman, a fireman, or a bandit. The employment man, on the other hand, may be inclined to underestimate the permanence of interest and to hire men on the basis of ability, disregarding interest entirely. This is probably unwise because the foregoing figures indicate some stability of interest.

This does not necessarily mean that the interest is inborn. We probably do have an innate interest in loud sounds, bright lights, and moving objects. Our interest in mechanical rather than literary pursuits is doubtless influenced by our experiences in childhood or later. An interest in chemistry or physics reflects to a still greater extent the environmental factor. The practical point is, however, that if a man approaches a job with a definite interest pro or con, the safest procedure is to assume that the interest will persist and it should, therefore, be reckoned with in occupational prognosis.

#### INTEREST AND ABILITY

There has been some discussion as to the relation between interest and ability. (593.) In one instance a group of students arranged the courses of their curriculum (mathematics, history, literature, science, music, drama, and hand work) in order of their interest and subsequently ranked these same subjects according to what they considered their own ability therein. The correlations between rank for interest and rank for estimated ability averaged

.89. The results were not so striking in another group of students when ranking for interest in college subjects was correlated with actual marks. It is rather probable, however, that accidental factors were involved in the academic grades and that the individual's estimates of their own ability came nearer to the real truth than did the grades obtained. This is further substantiated by the fact that estimates of ability and actual grades correlated to the extent of only .47. At any rate, there seems to be enough relation between interest and ability to be of some significance.

In the employment department of a Y.M.C.A. a group of men expressed their vocational interests. They were also given an intelligence test. The preferred vocations were located in the occupational hierarchy discussed in the previous chapter and the intelligence required by the job in which interest was expressed was correlated with the actual intelligence of the man expressing that interest. The coefficient was only .38. Making fairly liberal allowance, there were 36 per cent who possessed more intelligence than that required for the job in which they expressed interest, while 15 per cent had less than the requisite intelligence. The conclusion is drawn that the correlation between interest and ability is not over .50. (189.)

In the study of design engineers and sales engineers (*infra*), various special engineering aptitude tests were employed. Some interest questionnaires were also given. The interests correlated with the special capacity tests to the extent of .50. (394.)

The relation between ability and interest, then, is apparently not an extremely close one and there are, of course, obvious cases of lack of correspondence. A person may want to sing, but he may have a poor vocal apparatus; he may aspire to a berth on the police force, although he weighs only 110 pounds. Nevertheless, the studies just cited indicate a relation that is sufficiently close to merit some attention. It may be that one likes what he can do well. Or it may be that one devotes effort to the thing that he likes. In either instance interests are worth considering from the employment standpoint. If the former alternative is true, interest in a certain field would seem to indicate that the person had been successful in that general area and hence the interest might be

diagnostic of probable success in related fields. If the latter alternative is true, it indicates the desirability of employing a person for work in which he has some interest because he will then devote greater effort to it and use whatever ability he possesses.

#### METHODS OF MEASURING INTEREST

**Questionnaire.** Granted, then, that interests are of some importance to the employment psychologist, the question arises as to how this information regarding them may be best obtained. There are three general methods of approach to the problem of measuring interests; (1) by questionnaire; (2) by information tests; and (3) by more indirect methods. The questionnaire procedure consists essentially of asking the individual something about his interests or his likes and dislikes. For instance, he may be questioned regarding previous vocational interests, with a view to throwing light on his subsequent vocational interest. The following questions are typical:

1. Have you ever worked as a clerk in a store?.....
2. Have you ever conducted a house-to-house canvas?.....
3. What is the most responsible position you ever held?.....
4. What job that you have ever held did you like best?.....
5. Estimate how many hours during the past year you have spent working with tools, machinery, engines, and electrical apparatus.....
6. Have you ever constructed a piece of furniture or household appliance?.....
7. Did you construct it because you wanted the appliance or because you enjoyed making it?.....
8. Do you think you could find out what was wrong with a clock that would not run?.....; an electric motor?.....; an automobile engine?.....
9. Have you ever written a story that appeared in print?.....
10. Have you ever taught or tutored any one in a school subject?.....

Questions like these may be devised to cover a rather wide range of possible vocational interests. The selection of questions, of course, depends on the occupation for which they are to be used. If one is especially concerned with locating people who have been inclined toward social or mechanical vocations, the questions can concentrate particularly on these points.

## INTERESTS IN EMPLOYMENT PSYCHOLOGY 299

The questionnaire may also involve avocational interests. Consideration of these may throw light on tendencies that will be of later vocational significance. The following questions are typical:

1. What are your principal hobbies?.....
2. What is your customary recreation in the evening?.....
3. What sports or games do you like to watch?.....
4. What magazines do you read regularly?.....
5. What books that you read within the last year interested you most?..
6. Estimate how many hours during the past year you have spent in each of the following: driving an automobile.....; riding a motorcycle.....; horseback riding.....; hunting or rifle practice.....; swimming.....; tennis.....; golf.....; handball.....; other athletic sports.....
7. Assuming equal acting ability in each, which of the following do you prefer: dramas.....; musical comedy.....; vaudeville?.....
8. In which of the following activities have you ever taken part: dramatics.....; musical organizations.....; debating.....; politics.....; public speaking.....; reporting on a paper?.....
9. Have you ever made a collection of: stamps.....; coins.....; postal cards.....; what else?.....
10. In listening to radio what do you prefer: lectures.....; concerts.....; dance music.....; news items.....; logging stations?.....

Questions such as these may be devised to cover a great number of possible avocational interests which may be of vocational significance. The type of recreation one pursues may indicate his predilection for outdoor *vs.* indoor activity. His hobbies may give some clue to his inclination toward the mechanical. Miscellaneous activities will show something regarding propensity for literary, forensic, or physical work. For a particular occupation it may prove possible to determine the avocational interests that are of the greatest significance and devise questions to bring them out specifically.

There are some occupations that manifestly need an individual with social inclinations. Some of the questionnaire items may involve matters that will serve to indicate the social type of individual. A few typical questions follow.

1. At what age did you learn to dance?.....
2. Estimate how many smokers, lodge meetings, card parties, and other



- social affairs of your own sex you have attended during the past year. . . . .
3. Estimate how many mixed social affairs you have attended during the past year (include dance parties, socials, etc) . . . . .
  4. To what social clubs, fraternities, or business organizations do you belong? . . . . .
  5. What offices have you held in these organizations? . . . . .
  6. When single did you prefer to room alone or with a roommate? . . . . .
  7. Do you enjoy going to the theater by yourself? . . . . .
  8. How frequently do you play solitaire? . . . . .
  9. Have you a very few close friends? . . . . . a great many ordinary friends? . . . . . or both? . . . . .
  10. With how many persons do you maintain a social correspondence? . . . . .

Questions like these may be designed to bring out whether a person seems to enjoy the company of other people and to be more or less dependent on it or whether he is frequently satisfied to remain alone without social contact.

Instead of asking specific questions of the above sort, another somewhat similar approach consists of providing a lot of items which the subject marks according to whether he likes or dislikes them. For example, it is possible to consider the types of vocation in which the individual would be especially interested, providing he were not in his present vocation. Analysis of his likes and dislikes for various occupations may reveal certain definite trends in his interests. A typical set of questions of this sort is as follows:

Draw a circle around L if you would like doing that kind of work.

Draw a circle around D if you would dislike doing that kind of work.

Draw a circle around ? if you have no decided feelings toward that kind of work or know nothing about it.

Disregard any salary or social differences or any possible family objections. Consider only your own interest and satisfaction in doing each of the kinds of work listed. You are not asked whether you would take up the occupation permanently; you are merely asked if you would enjoy that kind of work. Assume that you have the ability necessary for each of the occupations.

Architect	L ? D
Automobile repairman	L ? D
Automobile salesman	L ? D
Bank Cashier	L ? D
Carpenter	L ? D
Draftsman	L ? D

## INTERESTS IN EMPLOYMENT PSYCHOLOGY 301

Editor of popular magazine	L ? D
Hotel keeper or owner	L ? D
Lawyer	L ? D
Machinist	L ? D
Newspaper reporter	L ? D
Pattern-maker	L ? D
Private secretary	L ? D
Purchasing agent	L ? D
Real estate agent	L ? D
Research worker in physics	L ? D
Stock broker	L ? D
Toolmaker	L ? D
U.S. Government astronomer	L ? D
Watchmaker	L ? D

In addition to obtaining information regarding a person's preference for occupations it is sometimes desirable to get his preference for various miscellaneous things. The following are typical of such a set of items:

After the following items draw a circle around L if you like the item, around D if you dislike it, and around ? if you have no particular feeling one way or the other.

Fat men	L ? D
Fat women	L ? D
Chinless people	L ? D
Energetic people	L ? D
Golf	L ? D
Hunting	L ? D
New Republic	L ? D
Movies	L ? D
Radio	L ? D
Crowds	L ? D
Fights	L ? D
Taking long walks	L ? D
Smokers	L ? D
Jokes on yourself	L ? D

The actual lists used would, of course, be much longer than the above and might be so selected as to sample a very wide range of interests.

**Information test.** Instead of relying on the subject's own statement regarding his interests or preferences which may be in some cases influenced by the use which he thinks is to be made of his statements or by his efforts to make such answers as will give favor-

able consideration to his case, it is possible to approach the matter more indirectly but more objectively. There is some ground for the assumption that if a person is interested in a certain field he will pick up information about it— will be more familiar with the terminology and with less obvious details that would presumably be overlooked by a person who lacked that interest. Consequently, an information test may give some indication of interest if the items are carefully selected. It is insufficient to ask questions which any one from casual observation would be able to answer. It is necessary to go further into details such as one would not encounter unless he had made definite effort to pursue the particular line under consideration. Below are given a few items from an interest test for agricultural engineers. In selecting items ordinary things that the students would meet in their everyday work in the college course were avoided. Technical journals were consulted and out-of-the-way things were selected on the theory that the student who was interested in this profession would naturally go beyond the ordinary required work of the classroom and would read additional things such as technical journals. In each item the subject checks the correct alternative.

1. Sunlight is an OXIDIZING AGENT: HUMIDIFIER: POISON:  
DISINFECTANT: TOXIN.
2. A track-laying tractor is used for ROAD-ROLLING: PAINTING:  
MILKING: LAYING TRACKS: PLOUGHING.
3. A surveyor's level is used for GRADING ROADS: FINDING AREAS:  
DETERMINING DIFFERENCES IN ELEVATION: MEASURING  
ANGLES: MEASURING PERPENDICULARS.
4. A pinion is a LOCK-NUT: SMALL GEAR: WHEEL: KEY:  
RACK.
5. A conveyor belt is used for moving GRAIN: GASOLINE: MO-  
LASSES: BARRELS OF SALT: LAMP CHIMNEYS.
6. Drain tiles are used for ELECTRICAL CONDUITS: BUILDING  
GARAGES: LOWERING SOIL WATER TABLE: PAVING:  
ROOFING.
7. The best blower belts are made of LEATHER: RUBBER: JUTE:  
SILK: HEMP.
8. Creosote is a FUNGICIDE: VARNISH: CATALYZING AGENT:  
BREAD FLOUR: SUGAR CLARIFIER.
9. Soil stack is a term used in BOILER FITTING: SURVEYING: CE-  
MENT MANUFACTURE: PLUMBING: SOIL ANALYSIS.
10. The best heat insulator is WATER JACKETING: SHEET CORK:  
HARD RUBBER: POWDERED ROSIN: CEMENT BLOCKS.

## INTERESTS IN EMPLOYMENT PSYCHOLOGY 303

To cite another instance of an information test used as an index of interests, a concern wished to know especially about the social interests of its applicants, whether they were "good mixers" and whether their interests led them into a wide social experience. Information questions were devised to cover a very considerable range of possible social interests. (472.) Only one item of each sort will be cited by way of illustration. The test involved items that were socially acceptable, that dealt with sports, and that were perhaps socially questionable. Some of the socially acceptable items are as follows. As in the preceding example the subject checks in each item the correct alternative.

1. Which of the following requires chairs? London Bridge ... Flying Dutchman ... Three Deep ... Going to Jerusalem ...
2. In what organization is 11 o'clock of special significance? Elks ... Odd Fellows ... Masons ... Knights of Columbus ...
3. In the song what follows the words "Blest be the tie that binds"? ... "Us in thy kingdom, Lord" ... "My faith on Calvary" ... "Loved ones of kindred minds" ... "Our hearts in Christian love" ...
4. What is a caucus? A national political convention ... An official county election ... A meeting of politicians within one party ... A secret political meeting in violation of the law ...
5. What is "French leave"? A dance ... Very few odds and ends left over ... Permission easily obtained ... Slipping away without notice ... Showing very polite manners ...

Some of the sport items are as follows:

1. What is the nickname of the Chicago Nationals? Cardinals ... Braves ... White Sox ... Cubs ...
2. Which of the following clubs has a wooden head? Cleek ... Brassy ... Niblick ... Mashie ...
3. What kind of a blow is a haymaker? Hook ... Uppercut ... Broad side ... Jab ...
4. What is the score when all 10 pins are knocked down? Strike ... Little slam ... Spare ... Break ...
5. What kind of a race is a derby? Trotting ... Pacing ... Running ... Hurdling ...

Some of the possibly questionable items are as follows:

1. How many spots on dice make a Little Joe? Three ... Four ... Seven ... Eleven ...

2. Which kind of wine is the strongest? Claret...Champagne... Sherry... Burgundy...Bordeaux...
3. What beats a flush? Fours...Straight...Three of a kind...Two pair...
4. What is the name applied to short lively chorus girls? Kittens... Ponies...Baby dolls...Footlight dodgers...
5. Which of the following is best for jazz dancing? Waltz...Fox trot ...Paul Jones...Minuet...

There is a possible error in the information test as a measure of interest, especially when dealing with items that are perhaps socially questionable. The subject taking the test may become suspicious as to its purpose. It is usually presented as a test of information, but the subject may "tumble" to the fact that information is being sought as to his actual social experience. He may realize that correctly answering questions regarding poker and the like will reveal the fact that he is familiar with the game and he may hesitate to commit himself for fear it will be held against him. This error is seldom involved when dealing with items about which no ethical question might be raised, but if some items like those in the foregoing test are used the results must be interpreted with care. It is probably advisable, furthermore, to put such a test at the very end of the program so that if an atmosphere of suspicion is developed, it will not affect the results of any other tests.

**Indirect methods.** Other methods of approaching interests are still more indirect. These methods are still in the experimental stage and too much stress should not be placed upon them until they are further validated. One of them involves what is ostensibly a memory test. The subject is given pairs of words and is required to associate the two words of each pair so that subsequently, when the first word of a pair is given, he can recall the second word that went with it. However, the pairs are chosen according to two principles. Some of them form perfectly ordinary associations such as "dog — cat," "brook — river," whereas others, perhaps alternate pairs, involve associations dealing specifically with the type of work in question. It is then probable that the person who is especially interested in the work will more readily associate the two words pertaining to it than will the person who is not. The

## INTERESTS IN EMPLOYMENT PSYCHOLOGY 305

following pairs of words are taken from such a test designed for agricultural engineers:

letter	stamp	brass	bearing
formula	equation	spider	spin
diamond	spade	level	transit
liquid	hydrometer	church	tower
watch	time	gasoline	kerosene
work	force	ocean	fish
rain	umbrella	engine	windmill

Alternate pairs, it is to be noted, deal with items familiar to agricultural engineers while the remaining pairs involve associations which are familiar to every one. The presumption is that persons with agricultural engineering interests will more readily associate "formula" and "equation," "liquid" and "hydrometer," etc., than will persons who do not possess such interests. If, however, the actual number of words recalled for the crucial pairs is taken as the final score, an error is introduced. One individual who has perhaps a very deep interest in this profession may make a low score on such words, not because of lack of interest, but because of poor memory, while another person with little interest but good memory may surpass him. This error may be obviated by taking the score on the crucial words relative to that on the normal words. The latter establishes the individual's general memory ability, and it is possible then to note by what per cent his performance on the crucial words exceeds or falls short of normal. If one individual does ten per cent better on the crucial words than on the normal and another does ten per cent worse on the crucial than he does on the normal, the former presumably has greater interest in the matter under consideration regardless of the intrinsic memory ability of the two individuals.

A different approach has been made with a sort of cancellation test. The subject is provided with a text containing irrelevant words, which are to be crossed out. In some instances the material is of an ordinary uninteresting sort. In other cases it is designed to appeal to some particular interest. The following example is a portion of such a test designed to locate persons who are ambitious and particularly interested in success and achievement:

## PART I

Advertising plays way to-day a very why conspicuous yet rôle in the yes management with of a business. It wan has assumed such proportions win in recent years war that it won is difficult to ton estimate the tan exact place tin which it occupies tip in commercial tub affairs. Over sib two thirds of the son cost of maintaining a see newspaper or sun magazine is derived saw from advertising say space.

## PART II

Suppose that gun it is success jot you want. There are few joys in this world that lab can compare lit with the joy of met achievement. Set your men mark and mat start climbing toward it. You mob will reach mud it if you keep mut at it. Be persistent pat and be pin patient. If you are in put Maine you can not wish rip yourself in rug California. But you'll sun get there sometime ton if you start tan and keep going tub even if you go rim on your hands tow and knees.

Part I, which is, of course, only a brief excerpt from the original test, is of an ordinary expository character with little appeal to any fundamental interest or tendency. Part II, however, is a "pep-talk" such as might appeal tremendously to a certain sort of individual. Some persons read all of this type of literature they can obtain and are much engrossed with the notion of personal success and "getting there." The theory of the test is that such persons will become so wrapped up in the passage while going through it that they will overlook many of the irrelevant words which they are supposed to cross out.

In this test, just as in the preceding, we must abstract from the individual's intrinsic ability in this particular sort of thing. One individual may naturally be less efficient than another in detecting irrelevant words or in speed of reading, and hence make a low score on Part II, not because of greater interest, but because of lack of ability of this sort. The uninteresting passage, however, serves to give an index of the individual's actual ability in this kind of performance. The results for Part II may then be taken relative to this, provided identical time limits are used in the two cases. The presumption is that the lower the score made by a subject on the "pep-talk" relative to the normal text, the greater was his interest in the passage.

EVALUATION OF MEASUREMENTS OF INTERESTS

**Technique.** The foregoing methods of measuring interests have been experimentally validated in a few instances. The statistical treatment of such results differs somewhat from that ordinarily used in connection with mental tests. In the former case it is a question of correlating with the criterion the total score of a test which comprises a considerable number of items. In the present instance it is not so much a question of total score as of the value of individual items. Inasmuch as we are investigating various interests, the main point is to determine which ones are significant in the particular problem under consideration. The usual technique consists of obtaining groups of people of different vocational status or of known differences of interest and then determining which particular items of the questionnaire or test are differential of these groups. If, for instance, a large per cent of a group of successful workers answers a particular item in a certain way while a small per cent of a group of unsuccessful workers does likewise, that type of answer is somewhat differential of the two groups. Or, again, we may assume that a group of salesmen have rather different interests from a group of engineers. If most of the salesmen give certain answers to items in an interest questionnaire, while few of the engineers give such answers, those items may be used to aid in differentiating sales and engineering interests.

When the difference between the proportions of the two groups giving a certain answer is not large, it is necessary to determine whether it is sufficiently so to be of practical value. This may be done by noting the proportion of those in each group who give the answer and of those who fail to give it and applying appropriate formulæ which give the "standard deviation of the difference."<sup>1</sup> If the actual difference is not at least twice the standard deviation,

<sup>1</sup> Such a formula is:

$$\frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}$$

where for instance  $p_1$  is the per cent of the successful salesmen expressing an interest in baseball,  $q_1$  the per cent failing to express an interest,  $p_2$  the proportion of unsuccessful salesmen expressing such an interest and  $q_2$  the proportion of unsuccessful salesmen failing to express such interest,  $N_1$  the number of successful salesmen and  $N_2$  the number of unsuccessful.



few statisticians would agree that it is of much practical significance. The difference might be obliterated by repeating the tabulation with other groups of subjects unless it is of at least this magnitude.

**Questionnaire results: sales vs. engineering types.** A group of students in a school for insurance salesmen, a group of design engineers and another of sales engineers were given, among other things, an elaborate interest questionnaire. (394.) One of the most differential parts of the questionnaire proved to be that dealing with preference for vocations other than the present one, in the fashion above illustrated. In this particular instance the subject marked each item merely plus or minus according to whether he would like or dislike working at that especial vocation. The items were tabulated then to show what per cent of each group marked a given item plus or minus. A few typical items from the original tabulation are given in Table XLI. This table shows the per cent

TABLE XLI. TYPICAL VOCATIONAL PREFERENCES OF DIFFERENT GROUPS  
Per cents expressing like (+) or dislike (-)

	INSURANCE SALESMEN		DESIGN ENGINEERS		SALES ENGINEERS	
	+	-	+	-	+	-
Architect.....	46	33	68	11	58	20
Automobile repairman.....	24	46	57	18	30	30
Automobile salesman.....	87	10	36	21	88	2
Bank cashier.....	56	26	29	46	38	40
Carpenter.....	16	48	39	18	18	44
Sculptor.....	18	42	18	54	6	44

<sup>1</sup> After Moore.

of each group marking each item plus or minus. For instance, "architect" is marked plus by 46 per cent of the insurance group and minus by 33 per cent. The difference is much greater for the design engineers, 68 per cent of whom mark it plus, while only 11 per cent mark it minus. An interest in architecture as a possible vocation seems more characteristic of engineers than

## INTERESTS IN EMPLOYMENT PSYCHOLOGY 309

of salesmen. The item bank cashier, on the other hand, shows the opposite trend. Insurance salesmen mark it plus much more frequently than they mark it minus, but both the engineering groups give a preponderance of minus marks. A similar analysis of all the vocations included in the list made it possible to select those which were most differential. The result is ten occupations which are chosen primarily by the salesmen type and ten which are chosen primarily by the engineering type. These are given in Table XLII.

**TABLE XLII. OCCUPATIONAL INTERESTS MOST DIFFERENTIAL OF ENGINEERING AND SALESMAN TYPES <sup>1</sup>**

<p><i>Engineering type:</i></p> <ul style="list-style-type: none"> <li>Architect</li> <li>Automobile repairman</li> <li>Carpenter</li> <li>Draftsman</li> <li>Government astronomer</li> <li>Machinist</li> <li>Pattern maker</li> <li>Research worker in physics</li> <li>Toolmaker</li> <li>Watchmaker</li> </ul>	<p><i>Salesman type:</i></p> <ul style="list-style-type: none"> <li>Automobile salesman</li> <li>Bank cashier</li> <li>Editor of popular magazine</li> <li>Hotel keeper or owner</li> <li>Lawyer</li> <li>Newspaper reporter</li> <li>Private secretary</li> <li>Purchasing agent</li> <li>Real estate agent</li> <li>Stock broker</li> </ul>
---	---

<sup>1</sup> After Moore.

If we compare the occupations in the two lists, the difference is rather obvious. Most of those chosen by the salesmen type are of the sort that involve social contacts, while those chosen by the engineering type are for the most part vocations in which one can work effectively by himself without very much personal contact. It seems that individuals in one group are somewhat inclined toward society and enjoy handling or motivating people, while the others are fundamentally inclined toward material objects rather than persons. After these lists had been determined, the individual blanks were gone over again considering only the differential items. In order to ascertain how well these two lists served to differentiate the salesmen from the engineers, the following procedure was adopted. The number of plus signs before vocations on the salesman list was added to the number of minus signs before vocations on the

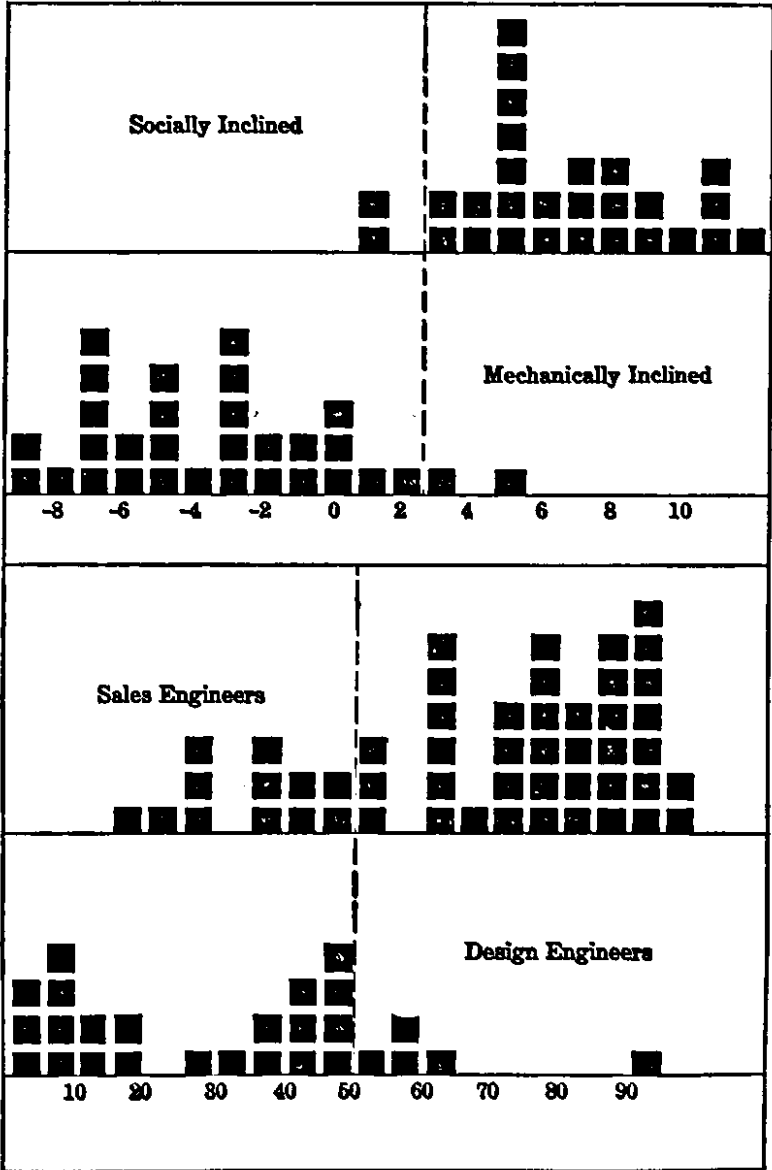


FIG. 8. INTERESTS OF OCCUPATIONAL GROUPS  
(After Freyd and Moore)

engineering list for a given individual in order to get his total in favor of sales. Conversely, the sum of the plus marks before engineering items and the minus marks before sales items gave his total in favor of engineering. These two totals were then added and his total in favor of sales reduced to a per cent of this sum. This indicated his per cent in favor of sales occupations. The average of these figures for all individuals amounted in the insurance group to 77 per cent, in the sales engineering group 68 per cent, and in the design engineering group 30 per cent. Thus it is evident that the interest expressed in these other vocations was rather differential of those specializing in sales as compared with those of more technical inclinations.

The results for the design and sales engineers are shown in the lower half of Figure 8. Along the base line are given the percentages which sales marks bear to the total marks. Each square indicates a person yielding the per cent that is given directly below that square on the base line. The upper group of squares is for the sales engineers and the lower group for the design engineers. The former are manifestly distributed more to the right, indicating a greater per cent of individuals marking the sales items. A critical score at 50 per cent makes a fair separation between the two groups. Very few of the design group exceed this score and only a small per cent of the sales group fall below it.

It was possible to follow up some of these sales and design engineering students after their graduation. Records were available of their later assignment in the engineering organizations for which they worked and of the success of their work. This later check-up shows that the original interest test would have correctly placed 85 per cent of them in the line of work at which they were ultimately proving successful.

In this same study other items of the questionnaire were similarly evaluated, but few of them seemed as differential as did the foregoing. The item referring to participation in a debating team showed that 32 per cent of the design group had done so, whereas the other groups contained approximately 50 per cent who had participated. Somewhat similar results held with reference to participation in dramatics.

*Social vs. mechanical type.* Another study along these same lines was conducted with college seniors in industries who were presumably mechanically inclined, in comparison with students in a school for insurance salesmen whose inclinations were supposedly more social in character. (177.) Among other things a questionnaire was given as to miscellaneous likes and dislikes and also as to likes and dislikes for various occupations. Differential items were determined in a manner similar to that used in the study just described. The items that proved most significant in differentiating the two groups are given in Table XLIII.

TABLE XLIII. DIFFERENTIAL ITEMS FOR SOCIALLY AND MECHANICALLY INCLINED <sup>1</sup>

Per cent checking symbols indicated

	SYMBOL	SALES	INDUSTRIES	WEIGHT
Actor.....	L	46	20	+1
Astronomer.....	L	13	36	-1
Automobile repairman.....	L	10	46	-1
Draftsman.....	L	10	43	-1
Factory manager.....	L	40	90	-1
Factory manager.....	D	33	3	+1
Locomotive engineer.....	L	20	43	-1
Machinist.....	L	10	50	-1
Machinist.....	D	63	26	+1
Magazine writer.....	L	53	23	+1
Magazine writer.....	D	20	50	-1
Preacher.....	L	23	3	+1
Ship officer.....	L	10	60	-1
Ship officer.....	D	56	20	+1
Shop foreman.....	L	3	56	-1
Shop foreman.....	D	73	16	+1
Specialty salesman.....	L	56	20	+1
Toolmaker.....	L	3	50	-1
Toolmaker.....	D	70	26	+1
Watchmaker.....	L	3	23	-1
Fat men.....	L	53	17	+1
Fat men.....	D	0	30	-1
Very polite people.....	L	47	90	-1
Very polite people.....	?	33	0	+1
Conventions.....	D	7	27	-1
Life.....	L	97	77	+1
Life.....	?	0	23	-1
<i>New Republic</i> .....	L	33	77	-1
<i>New Republic</i> .....	?	63	23	+1
Interviews.....	L	77	33	+1
Interviews.....	? or D	23	67	-1

<sup>1</sup> After Freyd.

The subject checked each item *L* if he liked it or would like to do that kind of work, *D* if he disliked it or would dislike that occupation, and ? if he had no decided feeling one way or the other. The table shows the per cent of each group checking the symbol indicated. For instance, 46 per cent of the sales group checked *L* for "Actor," while only 20 per cent of the industrial group did so. Such a check indicated to some extent a predilection toward salesmanship rather than industrial work and was accordingly given a weight of +1. An *L* for "Astronomer" was, on the other hand, checked more frequently by the industries group and indicated a tendency toward that kind of work. It was given a weight of -1. After these differential items had been determined, the original records for each subject were evaluated, considering only these items, and they were weighted according to the figures given in the last column of Table XLIII. These figures were then algebraically totaled to yield a composite score for each individual. The larger this positive score, the greater was supposed to be the tendency toward social rather than mechanical interests. The extent to which these composite scores differentiated the two groups may be seen in the upper half of Figure 8. Each square represents an individual making the composite score which is given directly below it on the base line. The upper group is socially inclined and the lower mechanically inclined. The separation of the two groups is pretty clear. If a critical score is taken between +2 and +3, there will be only two of the social group below it and only two of the mechanical group above it.

A very similar study was made comparing a group of seniors in industry and a group of mechanical engineering students. In this study it was possible in the same way to select a group of differential items that gave about the same sort of separation of the groups as in the present case.

**Information test results.** In addition to evaluating questionnaires as a measure of interests, the information type of test has been tried. The one described above for social interests was designed originally for use with salesmen. The complete test was given to a considerable number of salesmen. They were divided into five groups on the basis of their ability — highest success, high

success, fair success, doubtful, and inefficient. The information items were then tabulated to determine which were most differential of the groups. It proved possible in this way to select a fairly differential set. A poor score on the test served rather definitely to indicate a poor salesman, although a high score did not always insure a good salesman. Apparently a lack of the interests that were involved in the test tended to render one poor salesmanship material, but their presence was not of itself sufficient because other things not measured by this test were essential. A consideration of the most differential items reveals a tendency for the good salesman to be a man who has accepted social responsibility. One who is manifestly lacking in social experience seems to have poor chances for success in this line of work. (472.)

**Results with indirect methods.** The indirect methods of measuring interest have scarcely proceeded beyond the experimental stage. The method involving memory for word pairs above described was tried with a group of agricultural engineering students and the score on the test was correlated with an estimate by instructors as to "interest and industry." The test score — i.e., the memory for words related to agricultural engineering as compared to that for ordinary words — was found to correlate to the extent of .30 with estimated interest. The results were complicated by the fact that the test correlated with ability to an even greater extent, but there is some indication at least that there are possibilities in this method. The other indirect method above mentioned, namely, a cancellation test in which it was supposed that interest in the passage would detract from efficiency in cancelling irrelevant words, correlated  $-.30$  with estimated interest. This negative correlation is also in conformity with the theory of the test. More work must necessarily be done with the indirect methods before any great validity is attached to them, but they are cited simply to show the possibility of measuring interest in these rather indirect ways. (103.)

#### SUMMARY

Occupational success depends on other things than ability alone, and interest is one of them. There is some indication that interests

are rather permanent, and hence it is necessary to reckon with them in the employment situation rather than to rely on their changing to meet conditions. There is also some relation between interest and ability. Whether the ability motivates the interest or *vice versa* has not been determined. In either instance, however, interests are to some extent diagnostic of what the person will ultimately do in the occupation.

Several methods have been used to determine systematically a person's interests. A questionnaire may be devised dealing with previous occupational, with avocational, or with social interests, any of which may be of practical significance. Instead of answering questions the procedure is sometimes varied by having the subject check a list of items according to whether he likes or dislikes them. Information tests are sometimes used as a measure of interest on the theory that a person who is interested in a certain field will go out of his way to obtain more information about it and will remain "set" for anything pertaining to it, so that he will in the long run be able to give a better account of himself in an information test involving items in this field. Still more indirect methods have been attempted. In what is ostensibly a memory test, in which some items appealing to a certain interest are mingled with other normal items, it is assumed that relatively more of the former will be retained by a person with that particular interest. In a test involving cancellation of irrelevant words in a text, it is assumed that if the content of the text appeals especially to the person's interest he will become engrossed in it and mark relatively fewer of the irrelevant words.

These methods have been evaluated by administering the measurements or tests to certain occupational groups or to groups known to have some fundamental difference in interest and determining which items serve most clearly to differentiate the groups. It was possible from a list of items regarding which the subjects expressed their like or dislike to select a set which would differentiate fairly well the engineering type of individual from the salesman type. In quite similar fashion it proved possible to obtain differential items for the socially inclined as compared with the mechanically inclined. The information test as a measure of



interest proved of some value in discriminating different degrees of success in selling. There were indications that the successful salesman was an individual who had accepted social responsibilities. The more indirect methods of measuring interests have been tried to only a slight extent, but the correlations with estimated interest were somewhat encouraging. The whole matter of measuring interest and using such measurements in a practical way is still very much in the experimental stage, but satisfactory progress is being made.

## CHAPTER XII

### RATING SCALES

#### PURPOSE

**Study of non-measurable traits.** At the outset of the preceding chapter the point was made that capacity and ability do not constitute the whole story in predicting vocational aptitude. It is not merely a question of what the applicant *can* do, but of what he *will* do. He may be able to become a good calender man, but in actual practice he does not try to make the most of his opportunity to learn to operate the machine and so never succeeds. A man may have the requisite intelligence, or memory, or speed of reaction for a given job, but he may lack industry, initiative, tact, enthusiasm, persistence, or other traits or attitudes or tendencies that are needed to supplement his ability in order to make him a successful worker. In the present status of our science these tendencies or attitudes or traits or aspects of personality as distinguished from capacity or ability cannot be tested. The best that we can do is to obtain the judgment of persons familiar with the man in question.

Such traits cannot be stated in terms of items per minute, or some other quantitative unit, but they are nevertheless important in vocational prognosis. Things like initiative, tact, coöperativeness, leadership, or organizing ability, may be of outstanding importance in selecting men for, or promoting them to, positions of an executive, supervisory, or salesmanship nature in which personal contacts are paramount. The best procedure at present available for obtaining an indication of the degree to which such traits are present is the rating scale. Such scales are utilized in various ways by different organizations. Brief ones are used for estimating an applicant during an employment interview. Estimates of a systematic sort are obtained from previous employers, school teachers, or others who have been in touch with the applicant. Promotion from one department or job to another within an organization is a logical part of the employment program and it is for this purpose,

perhaps, that rating scales are at present most widely used. Many concerns have their employees rated periodically by their superiors. The results, on the one hand, indicate cases of maladjustment where transfer or special training is requisite, and, on the other hand, serve to locate promotional material. In many concerns vacancies are filled from within the same organization and the rating scale is useful in discovering the most promising individuals for promotion.

**More uniform method of expressing opinion.** Men are constantly observing one another and, from external behavior, inferring something regarding mental traits. These estimates are often built up almost unconsciously, but their effect is cumulative, and when a person is asked for an opinion regarding another he may realize that he actually has such an opinion already formed. These opinions, however, are of somewhat dubious value, especially in the form in which they are most frequently available. If a man is asked what he thinks of a given executive, or a given applicant, the answer usually involves some glittering generalities to the effect that he is a "good man," or a "poor man," or a "lazy worker," or "does not take hold." These terms are, of course, quite relative and mean radically different things to different persons. Being a good man in the estimation of one person may be equivalent to mediocrity in the estimation of another. General impressions of this sort are likewise apt to reflect prejudice. If the rater has had some unfortunate experience with the individual in question — for example, if he has encountered some single instance of carelessness — he is apt to impute the bad impression of this incident to the individual's entire personality. Hence it is desirable to abstract somewhat from these prejudices and general impressions and obtain the estimates in more scientific fashion. This can be accomplished to a certain extent by rating the traits separately and then combining them into a final rating. If, for instance, one is considering tact, initiative, and leadership separately, his judgment will probably to a lesser extent reflect his general impression or the influence of some single dramatic incident than if he is giving a single figure which is to evaluate the individual as a whole. Separate consideration of the traits in this manner obviates snap judg-

ment and insures that all the raters record their impressions in more systematic and above all in more uniform manner. The judges are somewhat less apt to disagree when rating traits separately than when estimating the worker as a whole, and if there is disagreement it is possible to analyze it because of the more uniform character of the whole technique.

**Educational value.** Another aim of the rating scale procedure in organizations where periodic ratings are made is to educate both the rater and the rated. The former comes to observe the latter more closely if he is required occasionally to rate him. In addition to arousing personal interest in the man it leads the rater to observe him with reference to different traits and consider them separately. The natural tendency is to devote attention primarily to the man as a whole or to some outstanding aspect. It is easy to dislike a man's face and overlook his other good qualities. The rating scale calls attention to these other qualities and teaches one to observe them too. One may discover that after all the man is rather skillful, ingenious, and coöperative. On the other hand, the scale may call attention to the man's laziness which had been previously overshadowed by his affability. In this way one's final opinion of the man and one's whole attitude toward him may be very appreciably changed. Furthermore, this procedure keeps the whole notion of personality alive in the mind of the executive.

The use of rating scales in an organization likewise has educative value for the employée who is rated. He realizes that he is being judged in essential traits. This may encourage a certain amount of self-analysis and evaluation and he may seek to determine his weak points with a view to improvement. He may also realize that the ratings have something to do with his status in the concern so that they serve to motivate him and prove an incentive to do as effective work as he can.

**Check on employees' progress.** Another purpose of the rating scale in some organizations is to give a periodic check on the employees' progress. Those whose development seems to be rapid and who are especially superior in certain traits may be considered as a source of supply for other higher positions within the organization and may be promoted. Others who seem weak in certain respects

may be transferred to other departments for which they are better adapted. These readjustments may become necessary because it is often impossible to apply the rating scale at the time of original employment. If personality traits could be measured by objective tests along with the capacities, persons could at the outset be placed in that line of work for which they are best fitted. As this is not the case and it is often necessary to wait until superiors are acquainted with the employees before estimates as to personality traits are available, these later adjustments are frequently desirable.

**Data to meet emergencies.** If an organization has ratings of its employees on file, they are often found useful in an emergency. Vacancies may occur unexpectedly and it may be desired to promote or transfer some one very quickly. The usual practice in such a case is to consult other members of the firm as to their general impression of certain possible substitutes for the position. If systematic rating scales are used at the time they can scarcely be properly evaluated. It is necessary first of all to train the raters. It is then desirable to rate a considerable number of employees of a given sort, compare the reliability of different raters, and if possible determine the validity of the ratings. It is often necessary to make corrections and run down special cases of discrepancy in which the rater apparently did not follow instructions. Only then can ratings yield their greatest value. If it is necessary to act quickly in an emergency, this careful procedure is not feasible. Consequently, it is more satisfactory if individuals who might be possible sources of supply for other positions or departments are rated systematically in advance and the final records filed for any contingency that may arise.

#### SELECTION OF TRAITS TO BE RATED

The mental characteristics to be included in a rating scale depend, of course, on the situation in which it is to be used. There is no one rating scale that is universally applicable any more than there is a universal test that can be used in selecting applicants for every job. The mental make-up of a successful salesman is considerably different from that of a successful executive. Consequently, if a

rating system is to be devised for salesmen, it should probably emphasize a different group of traits from that included in a similar system for executives.

**Traits that are present in varying degrees.** In selecting traits to include in a rating scale, there are certain ones which will be of rather dubious value. Such traits cannot be rated on a scale at all because they are not present in varying amounts. Such a thing, for instance, as loyalty is difficult to conceive in terms of more or less, for a person simply is loyal or is not loyal. The same thing may be said regarding honesty and various other traits in which it would be difficult to grade the individual on a scale. In such instances it is probably unwise to include the trait in the rating scale at all, because effort to estimate it in varying amounts will only be confusing. If such traits are significant, the rater can be required merely to check one of two alternatives according to whether the person is loyal or disloyal, honest or dishonest. The scale proper should comprise only traits that are present in varying degrees.

**Questionnaire.** In determining what traits to include in a particular rating scale the logical procedure is to consult persons who are familiar with the occupation in question. Members of the staff who have been concerned with employing or promoting certain kinds of employees have doubtless been using some personal unsystematic consideration of character traits such as appear in rating scales. It is possible then to circulate to such persons a questionnaire asking them to list or otherwise indicate all the traits which they consider important for this particular type of work. Certain traits on which they agree fairly well may then be considered of fundamental importance. As regards traits which are not so generally mentioned, they may be either discarded or made the subject of a conference which will bring out the reason why some members of the staff listed them while others did not. For instance, in devising a rating scale for salesmen each manager who was ultimately to use the scale was requested to submit independently a list of the traits which he ordinarily considered when estimating the value of a salesman. When the results were pooled the most frequently mentioned traits were included in the final

scale. They were as follows: experience, dominance, stamina, appearance and manner, enthusiasm, fluency, egotism, expansiveness. (278, 189.)

**Interview or conference.** Possibly a better procedure than the foregoing is to interview or otherwise confer with the members of the staff who are to suggest the important traits. In working with mere lists, unless elaborate definitions are used, there are apt to be ambiguities in terminology. A word may have quite different connotations for different individuals. If a man writes on his questionnaire that a job needs "coöperativeness," it is impossible to tell whether he means merely willingness to do what one is told or whether he considers further the tendency to anticipate the needs of others and to govern one's self accordingly in advance. The real meaning which he attaches to the term can be brought out in a personal interview. Sometimes this information is obtained in the interview conducted for the broader purpose of job analysis. (Cf. Chapter XV.) Sometimes it is desirable to have a conference as above suggested to iron out any apparent disagreements between the different members. It is well, however, to get each person to commit himself independently in the first place, because in a conference those who speak first may exercise a certain amount of suggestion upon the others, calling their attention to traits that had not occurred to them and to which they had originally attached little significance or minimizing the importance they had attached to other traits by failing to mention them. If each man's unbiased opinion is at the outset a matter of record, the conference is valuable in determining the reasons for various disagreements.

**Preliminary list from which to select.** It has sometimes facilitated the procedure of questionnaire or interview to provide in advance a fairly exhaustive list of traits from which the persons consulted may select those which they deem important. Persons who find it difficult to recall, when requested, the traits which they consider in evaluating their subordinates may find such a list helpful. Various lists and classifications of traits which might be valuable for this purpose are available. A typical one classifies a large number of traits under the following captions. (264.)

1. General intellectual: ability, alert, bright, intelligent, keen, thinker (good).
2. Special intellectual: breadth, scholarship, initiative, originality, good judgment, resourceful, mature mentally.
3. Efficiency of performance: accurate, capable, efficient, responsible, thorough, careful, expresses self well.
4. Efficiency in attitude: ambitious, diligent, determined, energetic, enthusiastic, persistent, prompt, industrious, painstaking, will succeed, willing to work.
5. Social — indicating control of others; executive ability, forceful, influential, leadership, inspires confidence.
6. Social — moral: character (strong), altruistic, conscientious, dependable, earnest, faithful, honest, loyal, reliable, steady, sincere, unselfish, high ideals, trustworthy.
7. Social — attitude toward others: adaptable, agreeable, charming, coöperative, friendly, genial, kindly, modest, independent, popular, social mixer, tactful, winning personality, poise.
8. Miscellaneous: appearance, habits, etc.

#### WEIGHTING THE TRAITS

Having determined the traits or qualities which are to be included in the rating scale, it is then essential to consider their relative importance, with a view to weighting them. It is quite possible that for salesmen tact may be twice as important as leadership. Just as when a number of tests are used for determining vocational aptitude, the predictive value is raised if the tests are properly weighted, so the value of a rating scale is increased if the proper significance is attached to each trait. While it would be possible in some cases to compare the ratings on different traits with a criterion, this is seldom done. Ratings are less objective, quantitative, and reliable than tests so that such procedure would scarcely be worth the effort. Moreover, the criterion itself is often a rating. The relative importance of the traits is generally determined in rather arbitrary fashion by using the best judgment of those who are familiar with the occupation in question.

**Frequency of mention in questionnaire.** If a questionnaire has been circulated or job analysis interviews have been conducted, it is possible to note how many times each item is mentioned in the questionnaire or in the interviews and to weight it accordingly. If, for instance, fifty people consider leadership important and



only twenty-five consider originality worth mentioning, leadership might receive a weight of 2, and originality a weight of 1.

**Pooled judgment.** Another possibility is to submit the final list of traits to a considerable number of judges and ask them to distribute 100 points among those traits. If, for instance, there are five traits involved, the judge is asked to assign a particular weight to each one such that the total of the weights will equal 100. These weights may be worked out in conference or made individually and the average weight assigned to each trait may be taken as approximately its final weight.

**According to reliability.** One other procedure for determining weights has occasionally been used. This consists of weighting the items roughly according to their reliability. The methods just described involve rather the validity of the traits, i.e., their relative merits in predicting some criterion. The present suggestion assumes that, since there is considerable difficulty in ascertaining the validity, it is better to look for the most reliable traits. If the judges agree with one another fairly well on some things and not on others, the former should receive more weight, not because they are more closely related to occupational proficiency, but because the ratings themselves come nearer to being a true index of the particular trait under consideration. It will be shown later in the chapter that some traits on the whole are estimated with greater reliability than are others. Those which tend to yield some objective product by which they can be judged, such as salary or bank account, are more reliable. Consequently, in lieu of actual measurements of reliability it is possible to attach more weight to such traits and less weight to those which are more subjective in character and hence presumably lower in reliability.

By way of illustration the weights assigned to the traits in a few rating scales will be cited. In the rating scale for salesmen mentioned above the following weights were adopted as a result of conference:

Experience.....	3	Enthusiasm.....	2
Dominance.....	3	Fluency.....	2
Stamina.....	2	Egotism.....	1
Appearance and manner.....	2	Expansiveness.....	1

In the army a rating scale for officers was devised. After a considerable amount of study and revision the final list of traits and their weights were as follows:

Physical qualities.....	15
Intelligence.....	15
Leadership.....	15
Personal qualities.....	15
General value to the service.....	40

In a large office force a rating scale for clerical workers was devised. The items were weighted by consultation with ten division heads of long experience. The items were weighted in two ways — one for clerical duties involving only individual work and the other for clerical duties where supervisory work is entailed. The qualities with their weights for individual and supervisory work follow:

	Individual	Supervisory
Appearance.....	10	10
Ability to learn.....	20	20
Accuracy.....	25	10
Dependability.....	10	10
Speed.....	20	5
Coöperativeness.....	7.5	10
Constructive thinking.....	7.5	10
Ability to direct work of others.....	0	25

**Incorporating weightings in the rating blank.** When the rating procedure is put into practical use, it may be arranged so that the rater gives it no concern whatever and the weighting is subsequently done by whoever evaluates the data, or the weighting may be actually embodied in the rating blank that is provided. In the case of salesmen just mentioned the rater estimated each trait in the same terms and on the same basis. When the results were totaled, the rating in experience was multiplied by 3, while that for stamina was multiplied by 2, etc. This was likewise the case with the scale for clerical workers just mentioned. All the workers were estimated similarly for each trait on a graphic scale (*infra*) which had the same maximum in each instance. If, then, the person was being considered for supervisory work, the different estimates were multiplied by one set of constants before being totaled, while if he

was being considered for individual work they were multiplied by the other set of constants. In the army rating scale, on the other hand, the rater used a master scale and considered his subordinates by comparison with other officers on the master scale. It was so arranged that in physical qualities he could assign a maximum value of 15, while in general value to the service he could assign a value up to 40. In this way the weighting was done in the actual process of rating, while in the other two cases it was done subsequently.

#### DEFINING THE TRAITS

**Avoid individual interpretations.** It is usually insufficient merely to present the name of a given trait and require a person to estimate somebody with reference thereto. If, for instance, the scale mentioned "executive ability," the rater might construe it either from the standpoint of planning or from the standpoint of ability to get things done. If a scale involved the term "originality," this might be interpreted either as ability to work without supervision or as actual inventive capacity. Hence it is necessary to define the traits in more detail. It may even be necessary to make tentative definitions and then revise them after preliminary use. In the army rating scale, for instance, the last item was originally "general value to the regiment." Later this became "general impression," and finally "general value to the service." Similarly, intelligence in the early form of the scale was described as "ease of learning, capacity to apply knowledge, ability to grasp and solve new problems." Later this definition became "accuracy, ease in learning, ability to grasp quickly the point of view of the commanding officer, to issue clear and intelligent orders, to estimate a new situation, and to arrive at a sensible decision in a crisis."

This illustrates the desirability of selecting the definitions rather carefully and of revising them if necessary in order that the persons using the rating scale will have in mind exactly the sort of thing that is desired. It has been suggested by some workers in this field that in the final form of a scale it may be better to omit altogether the actual name of the trait and to include simply the definition.

The theory is that if the trait is actually named, the rater may simply read the name and devote little attention to the detailed definition, thus putting his own interpretation on the name. Omitting the name compels him to read the definition and to approach more closely the trait which it is actually desired to have him rate.

**Objective preferable to subjective.** In defining traits it is further desirable to do so as far as possible in objective rather than in subjective terms. As mentioned above, objective traits which represent reactions to impersonal things or situations or tasks, and which tend to yield some objective observable product, are rated more reliably than the opposite type. Some traits may be defined in either objective or subjective forms, and in such a case the former is to be preferred because the raters, taking a more objective attitude, will tend to be more reliable. Consider for instance, leadership. An objective definition might be somewhat as follows: "Rate this executive in terms of the success which he has shown in developing a loyal and effective organization by administering justice, inspiring confidence, and winning the coöperation of his subordinates." Here attention is called to actual objective accomplishment, such as the organization he has developed and to the way his subordinates react as a result of his leadership. A subjective definition of the same trait might read thus: "Rate this executive's initiative, force, self-reliance, decisiveness, tact, ability to inspire men and to command their obedience, loyalty and coöperation." This definition calls attention merely to the subjective traits rather than to anything that results from their presence or absence. Again, personal appearance may be defined objectively; as, "consider how favorably he impresses people by his physique, bearing and manner"; and subjectively, as, "personal attractiveness, cleanliness, neatness, and dress." The presumption is that traits defined in the more objective manner will be more reliably rated.

**With reference to particular situation.** It is further desirable to define the traits with reference to the situation in which they are to be used. The definition that would be most satisfactory for rating an executive might differ somewhat from the definition that would

be most satisfactory for rating a subordinate. For instance, in a scale for executives, foremen, and supervisors, coöperativeness is defined as: "Success in winning the coöperation of his subordinates, in welding them into a loyal and effective working unit." In a scale for other workers in subordinate positions this same trait is defined as: "His attitude of helpfulness toward others, his inclination to coöperate in manner as well as in act with associates and superiors." Or again in the first scale initiative is defined as: "Success in doing things in new and better ways and in adopting improved methods in his own work." In the scale for workers it is defined as: "Success in going ahead with a task without being told every detail; ability to make practical suggestions for doing things in new and better ways." It is obvious that the definitions in the first scale stress different aspects from those in the second. It is hence necessary to consider definitions of the traits, as well as the actual traits themselves, with reference to the situation in which they are to be used, and to call attention to the particular things involved in that actual situation.

Where traits of rather general character are used, it is sometimes desirable, instead of defining them for the particular situation, to define them for a number of typical situations. Take for instance a trait like "self-assurance." It is possible to give a number of situations in which it might have opportunity to manifest itself and then to state for each a response that would apparently indicate a positive manifestation of this trait and another that would indicate a negative manifestation. (Cf. 322, 157.) The following situations are each followed by a possible positive and a possible negative response: (1) A new situation demanding response: positive — undertaking with readiness, carried out beyond demands; negative — excessive inquiry and waiting for directions. (2) Many tasks inviting response: positive — acceptance of many; negative — carrying a light load. (3) A task demanding preparation: positive — tendency to undertake without thorough preparation; negative — careful preparation. (4) Opinion asked: positive — readily given; negative — modestly withheld or qualified. (5) Contradicted when asserting one's own memory of an event: positive — denial of error; negative — acceding. The rater may

be required to estimate the person in each of these hypothetical situations or to consider them in making a single estimate for the trait in general. At any rate, calling attention to these situations clarifies the matter and is conducive to more accurate ratings.

By way of illustration one complete set of definitions will be given of the traits included in a rating scale for foremen used by a large paper manufacturing concern. (433.)

1. *Trade ability.* Kind and amount of trade experience, knowledge of and resourcefulness in using machines, tools, material, and trade methods.

2. *Ability to plan and supervise.* Ability to maintain standard quality work, to place help where they can do the best work, to plan ahead so as to have materials, men, and tools ready to get out orders on schedule time with minimum production cost and to keep a steady flow of work through the department.

3. *Ability to handle men.* Initiative, decisiveness, resourcefulness, energy, self-control, ability to see faults with his help, to earn their respect, goodwill, and confidence, to maintain just discipline and a stable working force.

4. *Ability to teach.* To explain work clearly to a beginner, to gain his confidence and make him interested in his work; success in developing all-round men, in bettering the men of lower grade and increasing generally the knowledge and skill of the help under him.

5. *General value.* Years of service, ability to understand and carry out the company's policies, orderliness in his department, readiness to cooperate in giving new ideas a fair trial.

#### MAN-TO-MAN RATING SCALE

**Construction of the master scale.** There are three types of rating scales that have been quite extensively used. The first of these involves man-to-man comparison. The classic example of this type is the officer's rating scale developed during the war. Its general principle involves comparing one man with another, rather than merely assigning him some number or letter. Its outstanding feature is the construction at the outset of a master scale. This consists, for each trait, of a number of individuals of the same type as those on whom the scale is to be ultimately used. These individuals are selected at the average and extremes of possession of the trait in question and their names written opposite appropriate

rating values. Then when a new man is to be rated he is actually compared with the men on the master scale and given a rating similar to the number assigned to the man on the scale whom he most resembles.

In the army scale, for instance, the rater was instructed to make out a list of some 12 to 25 officers of his own or lower rank with whom he was well acquainted and to include in the list men who were extremely good as well as extremely poor. He then considered the trait of "physical qualities," such as bearing, neatness, voice, energy, and endurance, and, disregarding every other characteristic, selected the officer who surpassed all others in this respect. This officer's name—e.g., Captain Smith—was then entered on the first line (cf. the typical master scale, *infra*) after the word "highest." Then he selected the one who showed the greatest lack of, or deficiency in, the physical qualities under consideration, and listed him on the line marked "lowest" (Lieutenant Briggs). Then he selected a third officer, about midway between these two, entering his name on the line marked "middle" (Lieutenant Brown). Two others were then chosen, one midway between the highest and the middle officer, and the other midway between the lowest and the middle officer. Weights for the five degrees of the trait had been previously determined. The highest degree had a weight of 15, the high 12, the middle 9, as shown in the blank. This then constituted the master scale for physical qualities. In the same manner the rating officer made out a master scale of five men for "intelligence," having in mind only this one trait while making out this scale. Similar master scales were then made for rating "leadership," "personal qualities," and "general value to the service." A typical master scale is as follows:

#### RATING SCALE FOR OFFICERS

##### 1. *Physical qualities*

Physique, bearing, neatness, voice, energy, and endurance.  
Consider how he impresses his men in the above respects.

Highest	Captain Smith.....	15
High	Lieutenant Jones.....	12
Middle	Lieutenant Brown.....	9
Low	Captain Doe.....	6
Lowest	Lieutenant Briggs.....	3

**II. Intelligence**

Accuracy, ease in learning, ability to grasp quickly the point of view of commanding officer, to issue clear and intelligent orders, to estimate a new situation, and to arrive at a sensible decision in a crisis.

Highest	Captain McCraig.....	15
High	Captain Miers.....	12
Middle	Lieutenant Graham.....	9
Low	Lieutenant Stewart.....	6
Lowest	Captain Hill.....	3

**III. Leadership**

Initiative, force, self-reliance, decisiveness, tact, ability to inspire men and to command their obedience, loyalty, and cooperation.

Highest	Lieutenant White.....	15
High	Captain Cornell.....	12
Middle	Captain Patterson.....	9
Low	Captain McCord.....	6
Lowest	Lieutenant Carver.....	3

**IV. Personal qualities**

Industry, dependability, loyalty, readiness to shoulder responsibility for his own acts, freedom from conceit and selfishness, readiness and ability to cooperate.

Highest	Lieutenant Miller.....	15
High	Captain Hopkins.....	12
Middle	Lieutenant Richards.....	9
Low	Lieutenant Hunter.....	6
Lowest	Lieutenant Taylor.....	3

**V. General value to the service**

His professional knowledge, skill, and experience; success as an administrator and instructor; ability to get results.

Highest	Captain Welsh.....	40
High	Lieutenant Donnelly.....	32
Middle	Captain Brooks.....	24
Low	Lieutenant Parks.....	16
Lowest	Lieutenant Sickson.....	8

The original blank provided the rating officer lacked, of course, the actual names of the captains and lieutenants. He wrote them in himself in the process of constructing his own master scale.

Use of the master scale. After the officer had filled out the names on the scale in this fashion, he could then use it for rating subordinates. If he was to rate Lieutenant Adams, he would



compare him with the five men indicated on the scale for physical qualities, and if, for instance, Adams seemed most similar to Lieutenant Brown he would receive a rating of 9. If he were somewhat inferior to Brown, but not so poor as Doe, he would be rated as 7 or 8. Adams would then be considered with reference to intelligence and the other traits and values assigned in exactly the same fashion. The total of these ratings represented his final standing and was used for various military purposes. Obviously a maximum rating of 100 points was possible.

Scales similar to the foregoing may be developed for any type of occupation where it seems desirable. The essential feature is the construction of a master scale, comprising names of workers of the given sort. The persons who are then to be rated are compared man-to-man with those on the master scale. The following is a typical scale for minor executives which proved useful in one organization.

#### RATING SCALE FOR MINOR EXECUTIVES

- I. Appearance and manner.** Ability to inspire confidence and respect through his appearance and manner.
- |              |    |
|--------------|----|
| Highest..... | 10 |
| High.....    | 8  |
| Middle.....  | 6  |
| Low.....     | 4  |
| Lowest.....  | 2  |
- II. Leadership.** Ability to elicit the coöperation of his colleagues and subordinates, to promote morale and to develop a loyal and efficient organization.
- |              |    |
|--------------|----|
| Highest..... | 20 |
| High.....    | 16 |
| Middle.....  | 12 |
| Low.....     | 8  |
| Lowest.....  | 4  |
- III. Organizing ability.** Ability to plan work wisely, to discriminate the relative importance of its different parts and to delegate its administration properly.
- |              |    |
|--------------|----|
| Highest..... | 20 |
| High.....    | 16 |
| Middle.....  | 12 |
| Low.....     | 8  |
| Lowest.....  | 4  |

IV. *Initiative.* Ability to get things done.

Highest.....	15
High.....	12
Middle.....	9
Low.....	6
Lowest.....	3

V. *Ability to develop men,* by teaching them about their work, arousing their interest in it, and stimulating their desire to progress.

Highest.....	15
High.....	12
Middle.....	9
Low.....	6
Lowest.....	3

VI. *General value to the concern.*

Highest.....	20
High.....	16
Middle.....	12
Low.....	8
Lowest.....	4

There are several advantages in this procedure of man-to-man rating. In the first place, it gets away from letter grades or "per cents," which to many persons are variously associated with school grades. If the raters were requested, for instance, to assign each subordinate some per cent between 0 and 100, they would be quite apt to think in terms of what the passing grade was in their school career. Some probably were accustomed to a passing grade of 50 and some to a grade of 70, and this would have the effect of sliding the "passable" workmen appreciably up or down the scale so that the results of different raters would not be very comparable. In the second place, the master scale is a relatively permanent measuring device. One would not use a cotton yardstick for accurate physical measurements because it might shrink overnight. But one's notion of a "75 per cent man" or a "B grade man" may shrink or stretch in similar fashion depending on such causal things as the time of day, the digestive condition of the rater, or some compliment or insult that he has recently received. The master scale, however, should not shrink. Comparing the physical qualities of a group of officers with Smith, Jones, Brown, Doe, and Briggs (who ranged from highest to lowest) to-day and making similar

comparisons next week should yield comparable results. For while a "grouch" might lower one's opinion of the group that was being rated, it would also lower his opinion of Smith, Jones, Brown, Doe, and Briggs. The ratings would all be relative to Smith, Jones, etc., regardless of the mood of the rater.

#### RATING BY DEFINED GROUPS

**Linear scale.** Another method which is frequently used involves, not direct man-to-man comparison, but rating persons relative to the other members of a definite group. This group is used as a standard. The scheme is quite similar to that discussed in Chapter VI in connection with estimates by superiors used as a criterion. In that case, however, we were concerned merely with a single estimate for each individual as to his ability in the job, whereas now it is a matter of evaluating separate traits. (Cf. 383, 574.) A typical blank for such a rating scale is given at the top of the next page.

The introductory statement is practically self-explanatory, but in actual practical use it is well to go over it with the persons who are to do the rating and insure that they understand what is required. The traits would, of course, be defined in detail either on the rating blank or on a separate sheet. Such ratings may be quantified by measuring the actual distance of the check marks in millimeters or some other convenient unit from the left edge of the left column. The larger number will then indicate a higher rating.

In the form just noted a different sheet is provided for each person that is to be rated. One person at a time is considered. The procedure may be varied by arranging it so there is one sheet for each trait, as follows:

#### ENERGY

	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Adams.....	.....	.....	.....	.....	.....
Andrews.....	.....	.....	.....	.....	.....
Briggs.....	.....	.....	.....	.....	.....

RATING SCALE FOR EXECUTIVES

Imagine all the executives of your acquaintance divided into five equal classes on the basis of their possession of each of the following traits, a highest fifth, a next highest fifth, a middle fifth, a next lowest fifth, and a lowest fifth. Now take the first man whom you are going to rate and, considering only his energy, compare him with these other executives. If you think, for instance, that he falls in the middle fifth, place a check on the line after "energy" in the column headed "middle fifth." If you think on the other hand that he is among the best 20 per cent in energy, check in the column at the extreme right. Furthermore, if, after you have located him in the proper column you consider that he stands relatively high or low in that particular fifth, indicate accordingly by placing your cross to the right or to the left. In other words, the farther to the right the cross is placed, the higher the degree to which the individual possesses the trait in question. Now take the same man and, considering him solely from the standpoint of initiative, compare him with the total group in that respect. Indicate your judgment by checking on the line for initiative in the same way. Proceed in the same fashion with the other traits.

	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Energy .....	.....	.....	.....	.....	.....
Initiative .....	.....	.....	.....	.....	.....
Leadership .....	.....	.....	.....	.....	.....
Tact .....	.....	.....	.....	.....	.....
Organizing ability ..	.....	.....	.....	.....	.....

In this case the rater considers one trait at a time and goes through all the men with reference to that trait before considering the other traits at all. This latter procedure is theoretically preferable to the former. There is always a danger of considering general impression rather than the specific trait in question. (Cf. discussion of the "halo" effect, *infra*.) In the former method when evaluating the same man with reference to the various traits in immediate

succession, there is a danger that opinion regarding initiative will be influenced by the rating for energy made a moment before. In the latter method there is less danger of this association of one trait with another, and the rater does not have to make such a special effort to abstract from other traits when considering a particular one. This method involves a little more preliminary clerical work in typing the names of persons who are to be rated. Some executives, moreover, may dislike the procedure of considering one trait at a time for all the men, simply because it is less natural and perhaps more difficult. However, the satisfactory administration of rating scales involves, as will be brought out later, some training of the raters and it is possible in this way to revise their habits as the method necessitates.

The division into five classes as in the above instance is not particularly essential. The following is another division that has been used. Part of the introductory statement which is similar to that used in the preceding case is omitted.

Check in one of the columns running from "very high" to "very low" to indicate the person's standing in the trait. Try to let the percentages guide you as to the number of check marks to place in each column.

## LEADERSHIP

	4 PER CENT VERY BAD	11 PER CENT BAD	21 PER CENT POOR	28 PER CENT AVERAGE	21 PER CENT GOOD	11 PER CENT VERY GOOD	4 PER CENT EXCEL- LENT
Adams.. .. .							
Andrews .. . . .							
Briggs.. .. .							

**Assignment to classes.** Some concerns have found rating scales of the foregoing sort too cumbersome and difficult for practical use. Instead of checking in the five columns and making fine gradations they merely have the rater assign a number from 1 to 5 to each man in each trait. (282, 283.) These numbers may be defined somewhat as follows: A central rating of 3 means that the em-

ployee meets reasonably satisfactorily the recognized departmental standards in respect to this trait. A 2 rating means that the employee is deficient enough in the trait under consideration so that he has had to be warned, criticized, or otherwise spoken to about it. A rating of 1 means that the employee is so seriously deficient in the trait that, if it is an important one, he is under consideration for transfer or dismissal. A rating of 4 means that the employee stands out above the general run of employees of the department in respect to this trait, while a rating of 5 means that the employee stands out so conspicuously from even the 4 men that he ought to be distinguished from them. A large bank uses a method like this for periodic rating of its employees, giving for each trait or characteristic a statement or set of questions calling attention to the main points and following this with the numbers 1 to 5. The rater simply rings one of these numbers. The following item is typical:

Consider how he applies himself to his work. Does he make his daily tasks his main concern? Does he give his best and continuous effort to his work? Is he earnest, persistent, or easily distracted? Does he stick with his work till it is cleaned up. Does he use his time and ability to good advantage? Or does he tend to do as little as he can to "get by"? Does he need constant, occasional, or no supervision in order to get his work done on time?

1            2            3            4            5

The significance of the numbers is as above described. Similar procedure is followed for other items such as regularity of attendance, special knowledge or skill, tact, coöperation, ability to learn, responsibility, and general suitability. One scale is used for rating managers, another for the higher grade clerical workers and a third for the machine operators. The items in the different scales of course overlap to quite an extent. With such a technique the actual rating and also the subsequent recording of the results is naturally more expeditious, but fine gradations of the estimates are impossible.

#### GRAPHIC RATING SCALE

**Superiority to other methods.** The two rating methods just described have certain shortcomings. The man-to-man scale proves

rather cumbersome. It takes considerable time and effort to make up the original master scales satisfactorily, and even then the actual process of comparing individuals with the men of the master scale is tedious. Unless the raters are thoroughly "sold" on the value of the proposition, they are not inclined to devote to it sufficient time and effort. Consequently, the results are unsatisfactory, for unless the master scale is made very carefully the rating program is damned from the start.

The method of using defined groups is less cumbersome, but in the linear form is almost too abstract for the average person who uses rating scales. It is a trifle difficult for the untrained to think in this fashion of the total range of a trait and to differentiate between the total range for "initiative," for instance, compared with the total range for "energy." Moreover, it is difficult to keep in mind the five or seven degrees of possession of a trait so that all persons will be judged on the same basis. If individuals are rated on different occasions they may unintentionally be rated according to a somewhat different standard. Merely assigning each individual a number indicating in which of the five classes he falls is simple enough, but finer gradations are frequently desired. The graphic rating scale has been devised to obviate some of these difficulties. It is much less cumbersome and more expeditious than the man-to-man scale because there is no master scale to construct. There is no necessity for carrying in mind standards as to total range or different degrees of a trait because these are all indicated by colorful descriptive adjectives or phrases. The rater, moreover, can make as fine judgments as he wishes.

**General nature of the scale.** The graphic rating scale involves the name or definition of the trait or both followed by a straight line a few inches long representing the distribution of the trait from maximum to minimum. Instead of simply marking off arbitrary points on this line, as in the method of defined groups, descriptive adjectives are placed along the line for the guidance of the rater. These adjectives range from those indicating a high degree of possession of the trait to those indicating a low degree. The rater checks at some point along this line as in the former method, but is guided by these descriptive adjectives. For instance, in rating a

person as to social attitude the line might be drawn with the following descriptive adjectives.

Constrained and formal	Slightly reserved	Meets one half-way	Cordial and informal	Extremely breezy and informal
------------------------	-------------------	--------------------	----------------------	-------------------------------

**Construction of the graphic scale.** The previous discussion of the selection, definition, and weighting of traits is applicable to the graphic scale technique. The selection of the descriptive adjectives or phrases, however, requires special consideration. For one thing care must be exercised regarding the extremes that are selected. Occasions arise in which one word has several opposites and it is necessary to determine which is to be used. The word "ambitious," for instance, might be opposed to "lazy" or to "indifferent." The phrase "good leader" might be contrasted, on the one hand, with "too frequent friction in his department" or, on the other hand, with "has to be led." In the one instance we are thinking of leadership especially from the standpoint of maintaining harmony and in the other from the standpoint of actually telling people what to do compared with being told what to do one's self. When extremes have been selected in this fashion, the intermediate phrases must then conform to the extremes. If, for instance, leadership is construed from the standpoint of harmony, the intermediate adjectives should deal with that general sphere, such as "obtains good coöperation" or "men dislike to work with him." In selecting extreme terms one should, moreover, avoid those that are so far from the average that they will never ordinarily be used at all. In rating ordinary workers there would probably be no place for the term "inventive genius," even though some aspect of originality was being considered.

Effort should be made to select words that are as concrete as possible. Terms such as "very," "good," or "highly" should be avoided. It is much better to get something which connotes a rather definite situation. If an individual is being rated on sense of humor, it would convey to the rater a much more definite notion to say "often has to have jokes explained to him" than to say "poor sense of humor."

There is no fixed rule as to the number of descriptive adjectives that should be used. In general practice from three to five seem



satisfactory. Three terms give opportunity for two extreme values and one intermediate or average value, while five terms facilitate slightly more detailed grading. Five terms are usually sufficient to give the rater an adequate notion of the distribution of the trait.

It is not always necessary that the adjectives be equally spaced along the line. In fact there are instances in which they ought to be unequally spaced because some adjacent pairs may actually describe individuals who are more similar than those described by other adjacent pairs. For instance, a set of four phrases for rating leadership might be distributed something as follows:

---

Inspiring leader	Handles men well	Men have little confidence in him	Continued friction with subordinates
------------------	------------------	-----------------------------------	--------------------------------------

---

In this case the two intermediate phrases are more closely related to the end phrase than to each other. The terms "inspiring leader" and "handles men well" are both positive in character and somewhat related, while the other two are similarly related being both somewhat negative. Hence the largest space is left at the middle of the line in accordance with the actual distribution of the trait in question.

It is sometimes deemed advisable in organizing the scale, if a considerable number of traits are to appear on a given sheet, to arrange them with the high extreme sometimes at the right and sometimes at the left. If this is not done and a person is rather superior in most traits, the rater is making his marks consistently along the right edge of the blank. Then if he comes to a trait in which the person is actually somewhat inferior, he is liable to continue the tendency to mark toward the right or at least the impulse to continue will appreciably bias his judgment of the inferior trait. This tendency to get a general impression and rate the person accordingly in most traits is the bugbear of rating technique, and this graphic method with all extremes at one end aids and abets this tendency, for the rater is inclined to make all his marks in one position on the blank. If the extreme values are staggered, it breaks up this tendency and makes him scrutinize each line a little more closely.

**Typical graphic scales.** A few graphic rating scales will be given by way of illustration. (525, 41.)

**GRAPHIC RATING SCALE FOR EXECUTIVES, DEPARTMENT HEADS, FOREMEN AND SUPERVISORS <sup>1</sup>**

Consider his success in winning confidence and respect through his appearance and manner.

Inspiring    Favorable    Indifferent    Unfavorable    Repellent

Consider his success in doing things in new and better ways and in adapting improved methods to his own work.

Highly constructive    Resourceful    Fairly progressive    Routine worker

Consider his success in winning the cooperation of his subordinates in welding them into a loyal and effective working unit.

Capable and forceful leader    Handles workers well    Fails to command confidence    Frequent friction in his department

Consider his success in organizing work of the department or unit, both by delegating authority wisely and by making certain that results are achieved.

Effective even under difficult circumstances    Effective under normal circumstances    Lacks planning ability    Inefficient

Consider his success in making his department or unit a smooth running part of the whole organization; his knowledge and appreciation of the problems of the department.

Exceptionally cooperative    Cooperative    Not helpful    Difficult to handle    Obstructionist

Consider his success in improving his subordinates by imparting information, creating interest, developing talent, and by arousing ambition.

Develops workers of high caliber    Develops workers satisfactorily    Neglects to develop workers    Discourages and misinforms workers

Consider his success in applying specialized knowledge in his particular field, whether by his own knowledge of ways and means or through his use of sources of information.

Expert    Competent    Uninformed    Neglects and misinterprets facts

<sup>1</sup> After Scott and Clothier.

**GRAPHIC RATING SCALE FOR INVESTIGATORS, SECRETARIES, SPECIAL WORKERS AND OTHERS NOT CHARGED WITH SUPERVISION <sup>1</sup>**

Consider the ease with which this employee is able to learn new methods; the ease with which he follows directions.

Very superior	Learns with ease	Ordinary	Slow to learn	Dull
---------------	------------------	----------	---------------	------

Consider the amount of work he accomplishes; the promptness with which he completes it.

Unusually high output	Satisfactory output	Only average	Limited output	Unsatisfactory output
-----------------------	---------------------	--------------	----------------	-----------------------

Consider the neatness and accuracy of his work and his ability constantly to maintain high workmanship in these respects.

Highest quality	Good quality	Mediocre	Careless	Makes many errors
-----------------	--------------	----------	----------	-------------------

Consider his energy and his application to the duties of his job day-in and day-out.

Very energetic	Industrious	Spasmodic or indifferent	Needs constant urging	Lazy
----------------	-------------	--------------------------	-----------------------	------

Consider his success in going ahead with a task without being told every detail; his ability to make practical suggestions for doing things in new and better ways.

Very original	Resourceful	Occasionally suggests	Routine worker	Needs constant supervision
---------------	-------------	-----------------------	----------------	----------------------------

Consider his attitude of helpfulness to others; his inclination to cooperate in manner as well as in act with associates and superiors.

Highly cooperative	Cooperative	Not helpful	Difficult to handle	Obstructionist
--------------------	-------------	-------------	---------------------	----------------

Consider his present knowledge of his work and of other work related to it.

Complete	Well informed	Moderate	Meager	Lacking
----------	---------------	----------	--------	---------

<sup>1</sup> After Scott and Clothier.

## RATING SCALES

343

### GRAPHIC RATING SCALE FOR CLERICAL WORKERS<sup>1</sup>

*Appearance.* Neatness of person and dress.

	Appropriate	Neat	Ordinary	Passable	Slovenly
--	-------------	------	----------	----------	----------

*Ability to learn.* Ease of learning new methods.

	Very quick	Catches on easily	Needs repeated instruction
--	------------	-------------------	----------------------------

*Accuracy.* Quality of work; freedom from errors.

	No errors	Very careful	Few errors	Careless	Many errors
--	-----------	--------------	------------	----------	-------------

*Dependability.* How well can he be relied on to work without supervision?

	Very reliable	Trustworthy	Usually reliable	Unreliable
--	---------------	-------------	------------------	------------

*Speed.* Amount of work accomplished.

	Very fast	Rapid	Moderate	Slow	Very slow
--	-----------	-------	----------	------	-----------

*Coöperativeness.* Ability to work with others.

	Coöperative	Falls in line	Difficult to handle	Obstructionist
--	-------------	---------------	---------------------	----------------

*Constructive Thinking.* Ability to grasp a situation and draw correct conclusions.

	Shows originality	Resourceful	Carries out suggestions	Needs detailed instruction
--	-------------------	-------------	-------------------------	----------------------------

*Ability to direct work of others.* Ability to direct and gain coöperation.

	Gets maximum efficiency	Directs work without friction	Secures limited coöperation	Wastes man-power	Antagonizes
--	-------------------------	-------------------------------	-----------------------------	------------------	-------------

The following are a few items from a graphic scale used for salesmen.<sup>2</sup> Particularly to be noted is the effort to deal specifically with what the salesman does rather than with abstract qualities.

Does he strike out for himself in locating prospects?

	Waits to be directed	Discovers some leads	Exceptional "nose" for prospects
--	----------------------	----------------------	----------------------------------

Does he impress people as being sincere?

	All he says taken at face value	Usually inspires confidence	Gives impression of bulldozing	Arouses suspicion
--	---------------------------------	-----------------------------	--------------------------------	-------------------

<sup>1</sup> After Bills.

<sup>2</sup> From Kenagy and Yoakum's *The Selecting and Training of Salesmen*, by permission of The McGraw-Hill Book Company, Inc., New York.

Does he put in full hours?	100 per cent attendance and punctuality	Commendable, better than the average	Satisfactory	Irregular	Very poor attendance
----------------------------	---	--------------------------------------	--------------	-----------	----------------------

Does he use good judgment in handling complicated situations?	Acknowledged blunderer	Makes an occasional error	Can be depended on to use good sense	Exceptionally clever in handling situations
---	------------------------	---------------------------	--------------------------------------	---

Does he dominate an interview?	Agrees with everything a prospect says	Easily thrown off the track	Usually guides conversation	Directs conversation; ready with a comeback	Completely dominates an interview
--------------------------------	--	-----------------------------	-----------------------------	---	-----------------------------------

How carefully does he study each prospect, his needs and attitude?	Has poorly considered plans	Has loose plans for prospects	Knows all that is readily available	Makes careful plans for big prospects	Goes deeply into every prospect's affairs
--	-----------------------------	-------------------------------	-------------------------------------	---------------------------------------	---

To show the possibilities of such a scale in an entirely different field a few items from a rating scale for teachers are given. (175.)

Is he self-conscious or self-possessed?

Painfully self-conscious and ill at ease	Frequently embarrassed or flustered	Self-conscious at all times	Usually unmoved by actions or remarks with reference to himself	Always at ease; self-possessed
--	-------------------------------------	-----------------------------	---	--------------------------------

Is he alert or absent-minded?

Always wide awake and alive to present situation	Usually has his wits about him	Fairly alert	Frequently becomes abstracted	Head in the clouds; preoccupied
--	--------------------------------	--------------	-------------------------------	---------------------------------

Does he display a sense of humor?

Sees funny side of everything	Usually sees the funny side of things	Slow in response to the comic	Often has to have jokes explained to him	Takes everything literally
-------------------------------	---------------------------------------	-------------------------------	--	----------------------------

How popular is he with his students and associates?

Arouses repulsion; detested	Disliked	Arouses neutral attitude	Liked	Popular favorite
-----------------------------	----------	--------------------------	-------	------------------

Is he prejudiced or fair-minded?

Partial and prejudiced; intolerant	Opinionated; has well-developed dislikes	Tries to be fair; usually just	Always impartial and fair-minded
------------------------------------	--	--------------------------------	----------------------------------

Scales like the foregoing have proved valuable in certain organizations. Just as with mental tests, however, there is no guarantee

that they will work in the original form in all concerns. They should be scrutinized by members of the staff to determine whether they will probably meet the needs of the particular situation. It is possible that some of the traits indicated will be of little importance and that others should be added. The foregoing scales, however, are typical and the methods used in their development can be used in any similar project.

**Scoring the blank.** The actual score represented by each mark on the rating blank is its distance from the right or left edge of the blank. This may be measured directly with a ruler in millimeters or some other small unit. A simpler procedure is to use a celluloid stencil ruled into 5 or 10 vertical columns with the width between extremes equal to the length of the lines used in the rating scale. Such a stencil may be placed over the blank and check marks read directly according to the column in which they appear. It is even possible in this way, if desirable, to weight the different traits while scoring them. If, for instance, one trait is to receive twice the weight of the other, the stencil for the former may comprise 10 columns and for the latter only 5. In this way if the columns are numbered from left to right a check mark near the extreme right will receive a rating of 10 in one case and 5 in the other. On the other hand, the same stencil may be used for all the traits, and then if they are to be weighted unequally the resulting numbers simply multiplied by the appropriate weight.

#### RELIABILITY OF RATINGS

**Conformity to normal distribution curve.** In dealing with ratings just as in dealing with tests it is desirable as far as possible to determine their reliability and their validity. Some notion as to their reliability may be obtained by making a distribution curve of the ratings assigned by a given person and noting whether the distribution is normal. (Cf. p. 162.) The presumption is that traits of this sort are distributed in about the same fashion as are the various mental capacities, and hence that correct ratings of these traits will yield a normal distribution. The expectation is, for instance, that executives who are fairly capable at developing subordinates will predominate, while as we go toward the extremes

of those who discourage and misinform their subordinates and those who develop men of exceptionally high caliber the numbers decrease. Hence, if the ratings made by a given person differ considerably from the normal type of distribution, we may suspect that something is the matter. If the curve is skewed in one direction or the other with a predominance of high or low ratings, it is probable that he is using too strict or too lenient a standard. If the curve is steep with very little scatter he is probably not making sufficiently fine distinctions between the men, and is not considering the whole range of the trait. If it is suspected that the rater is too strict or too lenient, it may be possible to have him rate a group that is known to be mediocre and see if he assigns them the same extreme values. In such instances as the foregoing it is well to confer with the rater and show him his tendencies. He may then after such a conference rerate the men and get perhaps something like a normal curve. It is rather possible that by this procedure more reliable ratings will be obtained.

These facts also suggest the possibility of correcting the original ratings statistically. The procedure discussed in Chapter VI for making heterogeneous criteria comparable is applicable in this connection. In that case, it will be recalled, the estimates made by each foreman were converted into terms of the total distribution of ratings made by that foreman. A distribution of his estimates was made, the standard deviation computed, and individual estimates were then reduced to terms of deviation from the average, divided by standard deviation. In the present instance, where the ratings are all on incommensurable traits and considerable unreliability is to be expected, there may be some doubt as to the value of such refined statistical procedure. A scheme that has sometimes been used consists of taking a considerable number, perhaps 50, ratings made by a given individual, arranging them in order from best to worst and calling the best 10 per cent A, the next 20 per cent B, the middle 40 per cent C, the next 20 per cent D, and the lowest 10 per cent E. Subsequent ratings made by this individual may then be converted into these same letters. Thus, after all, whether the rater takes a high or low standard those whom he puts relatively high will receive an "A" rating indicating desirable possession of

the trait, while those receiving relatively low ratings will receive a grade of "E." An illustration of this procedure is given in Table XLIV.

TABLE XLIV. CORRECTION OF RATINGS

ORIGINAL CLASS	NUMBER	CORRECTED CLASS	NUMBER
I	5	A	5
II	10	B	10
III	11		
IV	6	C	20
V	3		
VI	5		
VII	3	D	10
VIII	2		
IX	3		
X	2	E	5

An executive rated some of his subordinates on a graphic scale for a given trait. The results were scored with a 10-column stencil in which column I was the highest degree of possession of the trait and column X the lowest. The distribution for 50 of these employees is given in the table — 5 in class I, 10 in class II, etc. It is obvious that the distribution is skewed toward the upper end — that he is rating a great many men rather high. The presumption is that he is too lenient and that the greatest numbers should occur in classes V and VI rather than in II and III. The correction was arbitrarily made by calling the best 10 per cent — i.e., the best 5 men — A, the next 10 men B, the next 20 C, the next 10 D, and the lowest 5 E. These letters appear in the third column of the table opposite the appropriate numbers. This gives an approximately normal distribution of the grades. Henceforth men that the executive rated in class I were called A; those in class II, B; those in



classes III, IV, and V, C; those in VI, VII, and VIII, D; while those in IX and X were E grade. In this way allowance was made for the man's leniency so that only those whom he placed in a relatively high position received a high rating and those who fell in the middle of his range, even though he classed them toward the high end of the scale, received an average figure.

**Agreement of raters with each other; man-to-man scale.** A more direct approach to the reliability of rating scales may be made by noting the agreement of different raters with each other in estimating the same group of subordinates. Such an evaluation was made of the reliability of the officer's rating scale used during the war. (497.) A few typical results of official and experimental ratings will be cited. When 300 men who had been in an officer's training school together from two to three months made up master scales and rated one another there was marked disagreement in the standing of an army officer in the opinion of his fellow officers. The results for ten typical officers are given in Table XLV. One col-

TABLE XLV. VARIABILITY OF RATINGS MADE BY FELLOW OFFICERS WITH MAN-TO-MAN SCALE <sup>1</sup>

OFFICER	LOWEST RATING BY FELLOW OFFICER	HIGHEST RATING BY FELLOW OFFICER
A	52	80
B	38	67
C	66	92
D	36	73
E	53	87
F	48	83
G	43	77
H	43	71
I	39	75
J	32	65

<sup>1</sup> After Rugg.

umn gives the lowest rating each man was assigned by his fellows and the other column the highest rating which he was given. Officer A, for example, was rated as low as 52 points by one of his fellow officers and as high as 80 by another; B has ratings as low as 38 and as high as 67. This indicates that there is a considerable chance that an officer will be located at some distance from his true position.

In other groups, where the raters were given considerable training and discussion and then rated all of their fellows whom they felt competent to rate, the results were somewhat similar. Most of the individuals varied as much as thirty points in the ratings that they were given by the other members of the group. It was estimated that chances were not over four to one that any rating would be within fourteen points of the true rating.

Some light was thrown on these discrepancies by analyzing the master scales which the officers used in rating the men. It was found, for instance, that an officer who was placed at the top — i.e., given the value of 15 on one master scale for a particular trait — was frequently given a value of 12 on another master scale, sometimes assigned 9, occasionally as low as 6, and even in a few instances as low as 3. The fact that different officers showed such discrepancies in placing men on their master scales might indicate one cause for the low reliability of the ratings obtained by means of this scale. In this study not more than fifty to sixty per cent of the time did the officer on the master scale have about the same value on other master scales in which he appeared.

Agreement of raters with each other: graphic scale. A somewhat different result has been found when the reliability of the graphic rating scale has been studied in this manner. (437.) Results are available in which the same workmen were rated by two different foremen. The agreement between the two raters was computed by the usual correlation method. These coefficients for different pairs of foremen appear in Table XLVI. With the exception of foremen A and F there is a fairly high agreement between the different pairs. These two men were shown in other studies of their ratings to be rather inconsistent and when rating men on different occasions did not agree very well with themselves.

TABLE XLVI. CORRELATIONS BETWEEN RATINGS MADE BY PAIRS OF FOREMEN WITH GRAPHIC SCALE <sup>1</sup>

FOREMEN	CORRELATION
A and F	.33
H and D	.78
J and K	.82
L and M	.63
N and O	.80
N and P	.75
O and P	.84

<sup>1</sup> After Paterson.

These ratings were converted into five letter grades in the manner above suggested for correcting ratings. All cases in which a man was assigned a rating by two foremen were then considered with reference to whether the letters assigned were identical. The number of cases in which this was true was reduced to a per cent of the total cases. Similarly the per cent of cases was found in which there was disagreement of one letter step, e.g., a rating of "A" by one foreman and of "B" by another. In like manner was computed the per cent of cases with disagreement of two letter steps, such as "B" by one foreman and "D" by the other. These results are shown in Table XLVII for carpenters and tool designers. It will be seen that there is perfect agreement in both groups in

TABLE XLVII. PER CENT OF AGREEMENT OR DISAGREEMENT BETWEEN TWO FOREMEN IN RATING MEN WITH GRAPHIC SCALE <sup>1</sup>

	CARPENTERS	TOOL DESIGNERS
Perfect agreement.....	52	62
Disagreement of one letter step.....	44	34
Disagreement of two letter steps.....	4	4
Disagreement of three letter steps.....	0	0

<sup>1</sup> After Paterson.

over half the cases. Disagreement by one step is fairly frequent, but there is only four per cent of each group for whom there is disagreement of two steps and there is no greater disagreement than this. These results, like those in the preceding table, indicate a fairly high reliability of the ratings. At least the graphic scale shows manifestly more reliability in this sense as far as the results have gone than does the man-to-man scale.

Agreement of rater with himself. Another approach to the reliability of ratings may be made by considering the agreement of the rater with himself. It is possible to compare an individual's ratings on different occasions noting merely whether he assigns approximately the same average rating in each instance. This indicates whether he keeps about the same subjective standard. The results of such a study are shown in Table XLVIII. It merely

TABLE XLVIII. AVERAGE RATINGS ASSIGNED BY A FOREMAN IN SUCCESSIVE MONTHS <sup>1</sup>

FOREMAN	FIRST MONTH	SECOND MONTH	THIRD MONTH
A	59	61	61
B	52	51	53
C	52	51	52
D	46	47	49
E	54	59	60
F	50	58	52
G	39	46	45
H	46	46	55
I	55	44	46

<sup>1</sup> After Paterson.

gives the average of the ratings assigned by each foreman at different times a month apart. It is obvious that with the first few foremen in the list the average tendency remains fairly constant through the successive months, whereas with the last few, and

especially with foremen H and I, the average varies considerably. This indicates that some of them get their ratings rather definitely stabilized at the outset so that their standard changes little, whereas with others this is not the case. This same stabilizing tendency is perhaps more clearly shown in Table XLIX. This table gives the

TABLE XLIX. CORRELATIONS BETWEEN SUCCESSIVE RATINGS BY THE SAME FOREMAN WITH GRAPHIC SCALE <sup>1</sup>

FOREMAN	FIRST AND SECOND RATINGS	SECOND AND THIRD RATINGS
B.....	.91	.96
H.....	.88	.92
C.....	.85	.86
G.....	.84	.92
I.....	.84	.90
D.....	.82	.90
F.....	.62	.66
E.....	.60	.82
A.....	.52	.88
Average.....	.76	.87

<sup>1</sup> After Paterson.

correlations between the first and second ratings made by a given foreman and also between his second and third ratings. The first few men in the list obviously are very reliable even at the outset, whereas the last few in the list are not so reliable. These latter, however, improve very considerably so that there is a much higher agreement between their second and third ratings. The averages for the two columns show the general tendency for greater reliability to characterize the later ratings. This doubtless reflects the practice which the rater has had and substantiates the need, to be brought out presently, for giving raters definite training and practice.

The foregoing discussion indicates the methods in common use for determining the reliability of ratings. It further indicates the necessity for such determination because some scales appear much more reliable than others at least in the situations in which they have been used. Moreover some raters are more reliable than others. In any rating project effort should be made as soon as feasible to get some notion as to the reliability of the ratings given, for otherwise one may unwittingly be employing a rubber yardstick.

#### VALIDITY OF RATINGS

The validity of a rating is as important as its reliability, but it is usually more difficult to determine. As previously stated, validity means the extent to which a measure correlates with a criterion. In many instances the criterion itself is a rating of some sort so that correlation is impossible. It is often difficult to get a production criterion and this is particularly true of many occupations of an executive nature in which rating scales are especially used. There have been, however, a few studies of the validity of rating scales which may be cited by way of illustration.

**Army rating scale.** One of the items in the officer's rating scale used in the army was intelligence. Many of the officers who were rated for this trait also took the army intelligence test. It was then possible to compare the intelligence of a group of men as estimated by the scale with their intelligence as measured by the test. (497.) A typical group of officers was divided into five classes on the basis of rated intelligence and their ratings simply noted by a number from 1 to 5. They were likewise divided into five classes on the basis of measured intelligence. The lowest group in measured intelligence had an average rating of 2.8, the next lowest 3.1, the middle 3.4, the next highest 3.1, and the highest 4.2. There is some indication that those of higher actual intelligence as measured by the test are placed higher in the rating scale. This is clear with the extreme groups. The next lowest and the next highest in the test, however, have both the same average ratings, while the middle group is slightly superior to either of them.

When individual intelligence scores and ratings were correlated in fifteen different groups of officers the correlation coefficients averaged

less than .05. The officers making these ratings had had little experience and training. After they had received some further instruction in the technique, the correlations for nine different groups were .08, .08, .09, .11, .14, .15, .20, .21, and .23 with an average of .15. The training apparently produced a slight improvement, but not a very marked one.

If, however, the ratings made by several officers on the same man were pooled to get an average rating for that man and these average ratings in intelligence were then correlated with measured intelligence, the coefficients for three different groups were .48, .51, and .36. The pooled judgments are manifestly more valid than the individual judgments. The conclusion was drawn that "the averaging of three or four judgments would locate a person in his proper fifth of the scale."

It was not possible with the other items on the officer's rating scale to determine validity in this way because a criterion was not available. The general conclusion was drawn, however, that to be satisfactory a rating with such a scale should be the average of at least three independent ratings. In many other situations it has been found that a pooled judgment is much better than an individual one and this has been shown to be particularly true of man-to-man scales such as that just cited.

**Ratings of salesmen.** With the rating scale for salesmen mentioned above some of the items were validated by comparison with annual earnings. With the first item given in the above illustration, of those who were rated as having an "exceptional nose for prospects," 16 were very good salesmen — i.e., earning over \$5000; 10 were good, earning \$2000 to \$4000; 4 were mediocre (\$1000 to \$2000), and none poor. On the other hand, with those rated at the other end of the scale as having to "wait to be directed," the numbers in these same four salary groups were respectively 0, 2, 7, and 8. With another item dealing with how well the salesman studies his prospect (the last item in the above illustration), of those who were rated in the best 30 per cent on the scale 91 per cent were classed in the successful group, whereas only 57 per cent of the entire group were so classed.

A concern using rating scales should strive, where it is possible,

to make some determination of their validity. Where production or salary or some fairly objective criterion is available, this can readily be done. In some cases more indirect criteria are available,<sup>1</sup> such as membership in technical or other organizations, holding office therein or being listed in *Who's Who*. It is also possible to follow up the individuals after a period and compare their later success with their earlier ratings. Unfortunately, in many instances it is necessary to be content, for a time at least, with a study of the reliability of ratings with no consideration of the validity.

#### SOURCES OF ERROR IN RATING PROCEDURE

We have seen that rating scales often have considerable unreliability and it is in point to consider some of the sources of error in this procedure with a view to obviating or mitigating them where possible. We have already mentioned the locus of considerable error in the man-to-man scale, namely, in the construction of the master scale, and the need of training and especial care in the construction of this master scale was pointed out. There are, however, other factors that may introduce errors into any of the rating procedures discussed above.

**Comparative reliability of the estimates of different traits** is one of the factors that must be considered. It has been found that some traits are more difficult to estimate than are others. The results of two studies bring out the differences rather well. (246, 79.) In one instance twelve judges and in another five judges rated a group of individuals with reference to a considerable number of traits. The variability of the judges or their disagreement with one another was computed for each trait. The results are shown in Table L. Group I was rated by the twelve judges and Group II by the five judges. To make the two studies more comparable, the average disagreement of the judges on a given trait is taken as 100. Figures smaller than this indicate closer agreement and figures larger than this indicate greater disagreement, than average. The traits in the table are arranged roughly in order of closeness of agreement. If we consider the average column the traits may be grouped

<sup>1</sup> Cf. the discussion of miscellaneous criteria in Chapter VI.



TABLE L. AGREEMENT OF JUDGES IN ESTIMATING VARIOUS TRAITS<sup>1</sup>

TRAIT	GROUP I	GROUP II	AVERAGE	CLASSIFICATION
Efficiency.....	75	92	83	Close agree- ment. Average 88
Originality.....	95	77	86	
Perseverance.....	75	101	88	
Quickness.....	90	88	89	
Judgment.....	100	78	89	
Clearness.....	104	75	90	
Energy.....	75	109	91	
Will.....	85	98	91	
Mental balance.....	110	81	96	Fair agree- ment. Average 100
Breadth.....	100	92	96	
Leadership.....	90	103	96	
Intensity.....	85	113	99	
Reasonableness.....	115	86	100	
Independence.....	104	98	101	
Refinement.....	90	116	103	
Physical health.....	115	92	103	
Emotions.....	120	91	105	
Courage.....	100	119	109	
Unselfishness.....	115	106	110	
Integrity.....	104	130	117	
Coöperativeness.....	125	113	119	
Cheerfulness.....	130	112	121	
Kindliness.....	120	125	123	

<sup>1</sup> From Hollingworth's *Judging Human Character*, by permission of D. Appleton and Company, New York.

into the three classes indicated showing close, fair, and poor agreement. It is now interesting to analyze the traits especially in the two extreme classes and note any general differences. The most noticeable thing is that the "close-agreement" traits are somewhat more objective in character than are the "poor-agreement" traits. By "objective" is meant that they tend to yield some objective results or products such as inventions, books, positions, salary, bank account, property owned, and the like. A man's efficiency or originality or perseverance is apt to yield objective

products to a greater extent than is his integrity, coöperativeness, or kindness. The latter traits manifest themselves more in a social situation, and after they have been manifested there is nothing objective to show for it. The objective traits likewise more often involve reacting to things rather than to persons.

In another more exhaustive study of personality terms, especially those used in recommendations, somewhat similar results were found. (264.) Eighty terms are studied and classified according to the agreement between raters who used them. Somewhat the same trend is manifest. The alphabetical list of those on which there is greatest agreement in rating men begins with "ability, adaptable, breadth, dependable, diligent, expresses self well, hard-working, industrious," while the list on which there is least agreement begins with "alert, ambitious, coöperative, bright, character (strong), charming, cheerful, dignity." The traits in the former list are obviously more objective in the above-mentioned sense. Selecting the ten most objective traits, as far as could be judged, the average index of disagreement (per cent of "maximum random disagreement") was .55, whereas with the less objective traits it averaged about .70. This was with men rating men. With women rating women, on the other hand, the difference is negligible. This brings out the necessity in this whole procedure of evaluating rating scales, of taking account of sex differences.

In this same study the traits were evaluated according to the classification given at the outset of this chapter (p. 323). With men rating men the closest agreement was found for the classification "efficiency of performance" and the least agreement for "social attitude toward others." With women rating women both these classes had something like the average amount of agreement and the closest agreement was manifested for "general intellectual" and the least agreement for "special intellectual." Here again the sex difference is manifest.

The clearest point brought out in these studies is that, at least with men judging men, the more objective traits yield more reliable ratings. This notion of the objectivity of traits is then of importance in the construction and use of rating scales. It substantiates the point made earlier that the traits should be defined in objective

terms as far as possible. It further indicates the desirability of selecting for the scale traits that have this more objective character. This selection will tend to increase the value of the whole procedure because such traits seem to have greater intrinsic reliability.

**"Halo effect."** Another source of error that is very common in rating procedure has been called the "halo effect." (592.) This is the tendency to allow the general impression of the individual to color very markedly the evaluation of specific traits. If a man impresses us favorably either in a general way or by virtue of some particular aspect of personality, or perhaps by some happy incident in our contact with him on the golf links, we are prone to invest his personality with a halo which sheds a luster upon his various traits and leads us to overestimate the desirable and to underestimate the undesirable in his personality. Conversely, if our general impression is unfavorable, this is inclined to lead to underestimation of many of his desirable traits and *vice versa*. The conventional halo is of a favorable nature. The present one works both ways. If one is estimating an individual's height he is little influenced by prejudice or by general impression of that individual in other respects, but if it is a question of tact or industry or coöperativeness there is considerable danger of this error. For instance, a case is cited (497) of an army captain who was selected by twenty other officers for the low man on their master scale in such things as physical traits, intelligence, and leadership. As a matter of fact this man made the highest score on the intelligence test of any man in the unit and had previously held a Rhodes scholarship. Conference brought out the fact that he had an unpleasant personality and was hard to get along with. Consequently he was rated low in intelligence and physical qualities, although these had, of course, little relation to his particular personality tendency.

In another instance a group of officers carefully rated a large number of aviation cadets on the standard officer's rating scale. Correlations were computed to determine whether a cadet who was given a high rating, for instance, in intelligence was similarly rated in physical qualities and *vice versa*. If these correlations were large, it would indicate some presence of the halo — i.e., that the officer in rating one trait was considerably influenced by some general

impression of the cadet. As a matter of fact the following correlations were found: intelligence and physical characteristics .51, intelligence and leadership .58, intelligence and personal qualities .64. These are higher correlations than one would expect the actual theoretical relation between the traits to yield. Experimental studies of intelligence in comparison with various measures of physical qualities, such as stature, strength, and agility, have shown the relation to be slight. It is evident that the officers in making their ratings fell into this common error of the halo.

This is the same tendency noted earlier (p. 36), where it was found that in estimating traits from photographs high correlations existed between such traits as humor, perseverance, kindness, courage, and intelligence. A person who looked as if he possessed a high degree of one of these looked as if he possessed a high degree of the others.

A determination of the magnitude of the halo effect was made with the results of two teachers who had rated the same group of pupils in seven traits. (574.) For each pupil there was computed a composite rating of the seven traits — one composite rating for each teacher. These composites indicated, as it were, the teacher's general impression of the pupil, and the more closely any given trait correlated with the composite, the greater was the effect upon that trait of the halo of general impression. The ratings on a given trait made by the two teachers were then correlated to determine, for instance, how well they agreed in estimating honesty. Then, by the technique of partial correlation (cf. Chapter IX), this same correlation was determined with the effect of the two composite ratings constant. The extent to which this partial correlation was lower than the original correlation showed how much the halo had raised the intrinsic relation between the ratings of the two teachers. These two sets of correlations are shown in Table LI. For instance, the two teachers apparently correlated to the extent of .47, in estimating honesty, but the intrinsic relation between their ratings, abstracting from general impression, was only .19, representing a difference of .28. Similarly, with all of the other traits except cleanliness the partial correlation is lower than the original. The average of all the partial coefficients is .25 lower than the average of

TABLE LI. MAGNITUDE OF HALO EFFECT IN CORRELATIONS BETWEEN RATINGS BY TWO TEACHERS<sup>1</sup>

TRAIT	ORIGINAL CORRELATION	CORRELATION WITH GENERAL IMPRESSION ELIMINATED	DIFFERENCE
Honesty.....	.47	.19	.28
Obedience.....	.39	-.04	.43
Courtesy.....	.41	.11	.30
Orderliness.....	.19	.10	.09
Cleanliness.....	.47	.55	-.08
Sportsmanship.....	.36	.00	.36
Promptness.....	.45	.09	.36
Average.....			.25

<sup>1</sup> After Symonds.

the original and this indicates roughly the magnitude of the halo effect in this particular situation.

The method adopted in many rating scales of dealing with one trait at a time is designed among other things to obviate this halo effect. It aids the rater in abstracting from the other traits while evaluating a given one. If he rates a man in all of the traits in immediate succession, the effect of one is quite apt to influence another and the general impression to influence them all. If he rates all the men on a single trait before considering the next trait, he tends to take an attitude of comparing the men with one another in one respect rather than considering the same man simultaneously in all respects. Even then, however, this halo effect is often present. The effort to define the traits under consideration more carefully and to define them in objective terms will aid in directing the attention of the rater to the specific trait under consideration and away from general impression. In training raters particular stress must be laid on this halo error, for at best it is one of the most insidious difficulties in the rating scale technique.

**Length of acquaintance.** Another factor to be considered in

ratings is the length of acquaintance. Obviously, if a superior has known a subordinate only a few days he can give only a rather poor account of his various traits. On the other hand, it must not be assumed that the longer the acquaintance the better, because a number of factors enter after long acquaintance to introduce error in the results. A study that bears directly on this point was made with ratings of over 1000 public school teachers. There is every reason to suppose that the same factors involved here would apply equally to executives rating their subordinates. (296.) The most obvious tendency was to overrate persons who had been known longer. In "general efficiency," of those known less than one year only 10 per cent were rated excellent, of those known from one to seven years 47 per cent were rated as excellent, and of those known from 8 to 25 years 68 per cent were excellent. One possible explanation is, of course, that those who had been known many years actually had been teaching many years and had improved in efficiency as a result. However, other studies have shown that skill in teaching does not improve with experience to anything like the extent required to explain these results. Moreover, when the teachers are rated as to "physical efficiency" very much the same trend is found, and it is scarcely plausible that physical efficiency should improve in this fashion with age when dealing with adults.

The results can be explained satisfactorily as a result of the acquaintance factor. A supervisor would dislike to concede that the persons under him had not improved under his supervision and if he rated them on a par with the more recent ones this would be tantamount to such a concession. Again one is apt unconsciously to identify himself with the older subordinates because they are more similar to him in age and this will result in more favorable consideration for them. His own interests are apt to bias him in such identification. One supervisor who had previously been an athletic director gave as a reason for selecting a certain man as his best teacher the fact that he was a "he-man." Another supervisor who was a vigorous Sunday-School teacher selected a certain woman as her first choice because "she holds up high ideals before her pupils." Finally, with older subordinates, one gets adapted to them and to some of their weak points. Various mannerisms and

personality defects cease to attract attention so that ratings after long acquaintance are liable to be too high.

While these results were obtained in rating school teachers, the same reasoning would apply to executives or others rating their subordinates. The hesitation to concede that older employees had not profited by training under one, unconscious identification of the older with one's self, and adaptation to their weak points would operate in industry to introduce a similar error in ratings. It appears that knowing the subordinate too long decreases the critical value of judgments regarding him.

A somewhat similar situation was found in another instance when considering, not length of acquaintance, but degree of friendship. A group of persons rated one another in a number of traits and also as to their degree of friendship with the rater. (533.) It developed that there was a tendency to overestimate the good traits of one's friends. Those which were overestimated in this way were quickness, proficiency, memory, persistence, adaptability, leadership, and scholarship. In addition to the query, "How long have you known the applicant?" it would be well to add, "How well do you know the applicant?"

#### TRAINING RATERS

One of the most important aspects of the rating scale procedure is the training of the persons who are to do the rating, whatever particular form of scale is to be used. In this preliminary training there are a number of points that should be particularly stressed and effort should be made to impress them upon prospective raters.

**Attitude.** One of these is the attitude with which the rater approaches his task. This should be objective and impartial. He must rate his friends on the same basis as other subordinates with whom he has only business contact. One has merely to listen to two women discussing the merits of their children to appreciate the danger of taking a partial attitude in making estimates. No effort should be made to cover up a person's weak points for if they are brought to light proper adjustments are often possible. Conscious prejudice sometimes is involved, but of more frequent occurrence is

some unintentional bias due to especial affability of the person rated or to some single incident of favorable or unfavorable character. It is a trifle difficult to give a poor rating to a man who is the "life of the party" or to give a high rating to one who has insulted you. It is important to teach the rater to abstract from all such things, to hold the individual, as it were, at arm's length and estimate him objectively and impartially.

**Basis for rating.** Consideration must also be given to the basis on which the rater is to make his judgments. It is advisable for him to base his estimate on actual rather than expected performance. The latter sort of estimate becomes more subjective and involves not only the rater's ability to estimate traits as he can judge from what he observes, but also his ability to infer therefrom how the person will behave at some future time. This is manifestly more precarious. Moreover, he must have under consideration only the present group of employees if he is rating them with reference to one another. Messengers obviously should not be compared with typists. The ratings should be made with reference to the particular kind of work that is involved or the special industrial situation under consideration. Initiative in golf and in the cost department may be entirely different things. A man may be energetic in collecting stamps, but lazy in figuring time slips. Patience in watching a cut with a machine tool does not necessarily reflect patience with one's family and *vice versa*. The rater should then be taught to consider the traits of the man on the job rather than the man at home or elsewhere.

**Standards.** The rater obviously has to judge according to some standard, whatever the particular technique used. As previously mentioned, some may adopt a standard that is too lenient and others one that is too severe. This may usually be ascertained from a distribution curve of the ratings made by a given man. If he places most people too high or too low this should be pointed out to him in conference and he should be required to justify certain cases if he still maintains that his estimate is correct. He should be told at the outset that the persons below and above average are usually fewer in number than are average persons because they constitute exceptions to the general rule. Frequently when a



rater's tendency to overrate or underrate is pointed out to him he will revise his ratings and henceforth use a more normal standard.

Once a standard has been adopted by a rater he should make every effort to maintain it constantly throughout the procedure. There is danger of relaxing or otherwise changing the standard in the course of time. The man-to-man scale was devised in the light of this very fact. With other types of scale it is possible to maintain the same standard throughout after adequate training and practice. It is often well to recur occasionally to some of the ratings made a little earlier and see if they still seem correct. If they do this will indicate that the same subjective standard is still being maintained.

Effort should be made as described earlier to distribute the ratings over a normal range rather than to bunch them. Some raters are afraid of making invidious distinctions and as a result give almost the same ratings to every one. They should of course have this called to their attention and be taught to distribute their ratings more widely. Another common tendency is to use greater care in making distinctions at the lower end of the scale than at the upper. Some raters will bestow the better estimates rather indiscriminately, although they take plenty of pains with the poorer ones. The fine distinctions are often just as important vocationally at the upper end for determining promotional material as at the lower end for detecting misfits, and the rater should learn to govern himself accordingly.

**Process of rating.** The essential aspects of the actual process of rating have already been brought out, but the rater should be watched to insure that he forms the habit of observing them. The ratings must be made independently. It is tempting to talk them over with others who are making similar ratings. If a colleague glances at one's ratings and makes some casual remark, one is tempted to reconsider and perhaps to make some compromise. If the colleague is to be involved the proper thing is for him to make similar ratings independently and then to compare the results statistically. It has been shown in various connections that greater validity is obtained by averaging independent estimates than by

having the judges sit together as a committee and make a joint estimate.

The other aspect of the process of rating that is essential to the success of most scales is judging one trait at a time. It is tempting for the rater to take one individual and consider him throughout. This process is often the more expeditious. He should be shown, however, the danger of the halo effect and convinced of the desirability of employing the other method.

**Sufficient time.** It is especially essential in training raters to convince them of the necessity of taking plenty of time. A busy executive who is accustomed to make quick decisions regarding matters of routine often finds it difficult or unpleasant to slow down and give the careful consideration to particular traits that is necessary. Consequently he must be "sold" on the value of the whole procedure so that, whatever the amount of time necessary for rating and rerating, he will be willing to devote that amount of time to the project.

**Conference.** Finally to safeguard the whole procedure frequent conferences should be held between the one in charge of the project and the persons making the ratings. It is insufficient to give the raters some printed directions and blanks and turn them loose. After they have had an opportunity to study the manual of directions it is a good plan to have a conference of all of the men and talk it over. Any difficulties that have occurred to them can be clarified on the spot. Many of the things mentioned above in this section can be explained to them and emphasized although subsequent repetition will of course be necessary. After this each one may well be asked to make out a sample set of ratings. These can then be reviewed carefully and criticized in the light of the foregoing considerations. Ratings by different men may also to advantage be compared to find those who agree and those whose ratings seem typical. When any shortcomings of a man's ratings appear his attention may be called to the fact. He can then rerate the same group or make other new ratings to see if he can profit by his previous mistakes. His second series of ratings may be similarly criticized and analyzed and perhaps compared with the first set and this procedure repeated as often as necessary. In a large banking

organization each rater has his ratings reviewed in personal conference three successive times and this procedure is repeated twice a year if necessary. (282.)

This training of the rater tends to make his results more reliable. This has been shown statistically, as for instance in experiments with the officer's rating scale where a group of officers after training provided better estimates of intelligence than they did before instruction. As previously mentioned the combined results of several raters are usually better than the results of one. A minimum of three independent ratings has been recommended as a result of statistical studies. If then the rating scale has been properly constructed, if the raters have received adequate training and if at least three raters make their estimates independently and the results are pooled the results will be found of value in many practical situations.

#### SUMMARY

Rating scales are necessary in evaluating various traits that are of vocational significance but that cannot be measured objectively. They are used by interviewers, by previous employers or acquaintances, with a view to initial employment and by executives and foremen with a view to promotion or transfer. They afford a more uniform method of expressing opinion regarding prospective or present employees as they deal less with general impression or prejudice and more with specific traits. They educate the rater in leading him to make closer observations of his subordinates and in keeping the notion of personality before him and they educate the employee who is rated in observing himself more critically. They often afford a valuable check on the progress of employees and if ratings are on file they afford data to meet emergencies such as could not be obtained in systematic and reliable form on short notice.

In selecting the traits to embody in a rating scale for a particular situation it is desirable to eliminate those that are merely present or absent and not present in varying degrees. The best traits may be determined by circulating a questionnaire to persons familiar with the occupation asking them to indicate those which they consider

most important. The most frequently indicated traits may then be included in the scale. A better procedure is to determine the traits in an interview or conference where ambiguities in terminology can be cleared up.

The next step is to weight the traits according to their relative importance. The frequency with which a trait is mentioned in the questionnaire or interview gives some notion as to its importance. The final list may be re-submitted to the executives with the request that they distribute a certain number of points among the traits and the average value assigned any trait may be taken as its approximate weight. In some cases the more reliable traits have been assigned greater weights not because the estimates are more closely related to the criterion but because they are truer indications of the trait under consideration. The weighting may be actually incorporated in the rating blank or the traits may be all rated on the same basis and the weighting done subsequently.

It is necessary to define the traits in order to prevent the rater from putting his individual interpretation on a term. It is better to define in objective rather than in subjective terms because objective estimates have greater reliability than subjective. It is also desirable to make the definitions with reference to the particular situation in which the scale is to be used.

The man-to-man rating method involves the construction of a master scale for each trait. This consists of the names of individuals who possess the trait in question to various degrees. Their names are written on the blank opposite appropriate rating values that have been previously determined. In rating persons they are compared man-to-man with the individuals on the master scale and given a rating similar to the number assigned the man on the scale whom they most resemble. The typical instance of the man-to-man scale is the officer's scale used during the war. The method, however, has been adapted to rating various other occupational groups such as executives.

Another method involves rating the individual relative to other members of a defined group. The rater may be required to imagine all the persons he knows engaged in the occupation in question divided into five classes of equal ability and then locate the given

individual with reference to these five classes. The blank may be presented in the form of a linear scale with the groups indicated by columns so that the rater can judge as finely as he wishes. A cruder scheme involves merely assigning each individual a particular number from 1 to 5, these numbers having been previously defined.

The graphic rating scale involves the name and definition of a trait followed by a line along which the rater checks at some point. He is guided by descriptive adjectives or phrases distributed along the line ranging from low degree of the trait to high degree. Care must be exercised in the selection of these adjectives or phrases so as to insure that the extreme ones are actually opposite and that the intermediate ones conform to the extremes. They should be spaced in accordance with the actual distribution of the trait and should perhaps be staggered with the high degree sometimes at the right and sometimes at the left, lest the rater drop into the error of making all his marks in about the same position. Graphic scales have been devised for many occupations such as executives, secretaries, clerical workers, and salesmen. The ratings can be quantified by measuring the distance of the check mark from one edge or by the use of a stencil ruled in columns.

The reliability of a rating scale ought to be investigated before it is put into any very general use. Some notion of its reliability may be obtained by determining whether the ratings made by a person conform roughly to a normal distribution curve. If the curve is skewed toward the high or low end, or is very steep and narrow, it indicates that the rater is setting too strict or too lenient a standard or that he is failing to consider the whole range of the trait. It is often necessary to correct the original ratings in the light of this fact and to consider as high only those rated relatively high and *vice versa*. Reliability may be further studied by noting the agreement of raters with each other. With the army scale there was a rather small agreement of different officers in rating the same men. These discrepancies appeared to a considerable extent to be due to the construction of the master scales. With the graphic scale more encouraging results have been found. Different foremen rating the same subordinates agreed rather closely in most instances. A

further indication of reliability is given by comparing successive ratings by the same man. His average ratings on successive occasions will show whether he is maintaining approximately the same standard. With the graphic scale rather high correlations were found between foremen's first and second ratings of the same men and higher correlations still between their second and third ratings.

The validity of ratings should be ascertained where possible, but often no criterion is available whereby to determine this validity. Estimates of intelligence in the army scale showed some relation to intelligence as measured by a test, especially if the ratings by three or more judges were averaged. Some items in a graphic rating scale for salesmen made some differentiation between those in different salary groups.

Certain sources of error in rating procedure may be noted. Traits that are subjective in character have appreciably less reliability than those that are more objective and that yield some products by which they may be judged. The halo effect is a particularly insidious source of error. This is the tendency to get a general impression of the individual and to rate him accordingly in all traits rather than to discriminate the separate traits. It can be shown in many instances that estimates of different traits intercorrelate more highly than they ought to. The length of acquaintance with the person who is rated is of interest. If it has been long the rater is apt to give too favorable an estimate due to unconscious identification of the older subordinates with himself, hesitation to concede that his long influence has not improved them and adaptation to their weak points.

Finally, the raters ought to receive systematic training. They must be taught to take an impersonal, impartial attitude; to rate the subordinate on actual rather than on expected performance and on performance in the special industrial situation under consideration. They must adopt a normal rather than an extreme standard as a basis for judgment and must then maintain it throughout. The actual process of rating should be carried through independently and one trait at a time should preferably be considered for the entire group. The rater must be convinced of the importance of devoting ample time to the project. To safeguard the whole

procedure, frequent conferences should be held to review the ratings with those who made them and to discuss any errors that are manifest.

When the rating scale has been properly made and at least three trained raters make independent judgments of a group of individuals the combined results will be of some value in the practical situation.

## CHAPTER XIII

### MISCELLANEOUS DETERMINANTS OF VOCATIONAL APTITUDE

#### VALUE

EMPLOYMENT psychologists have devoted most of their efforts to the use of mental tests of one sort or another for the prediction of vocational aptitude. This is due considerably to the fact that the tests are objective and yield results that do not depend on the judgment of the applicant or of persons familiar with him. The tests, moreover, are quantitative and usually yield a wide range of scores. All these things contribute to the reliability and validity of the results.

**Supplement tests.** Granted that test procedure is generally superior to less quantitative or objective methods, there is nevertheless the possibility that these latter may be valuable as a supplement to the tests or even in lieu of them in instances where tests are not feasible. With reference to the former possibility we have previously seen that, in deriving a regression equation for predicting vocational aptitude, the more variables evaluated, the greater the chance of finding a group which, if properly weighted, will give a high correlation with the criterion. With the average marksman a shotgun is more effective than a rifle. So with a group of tests or other measurements none of which can give a perfect vocational prediction, the more that are tried, the greater the chance of finding some that are valuable for the purpose at hand. In the discussion of weighting a group of vocational tests, it was suggested that it is advisable to try out a rather wide range of tests and select for further careful study those which have high correlations with the criterion and low correlations with each other. It often develops in an employment research that most of the tests used intercorrelate rather highly. Hence there is the possibility of turning to other variables besides tests — e.g., such things as items of personal history — which may perhaps show some correlation



with the criterion and likewise a low correlation with the tests. If the correlations with the criterion are sufficiently large and those with the tests sufficiently small, the addition of these variables will then increase appreciably the validity of the whole procedure of prediction. At any rate, it has seemed worth while in many instances to determine whether there are any additional variables of this sort available and to evaluate them at least in a rough statistical fashion with a view to further refinement of treatment, providing they are promising. It is quite possible that tests plus certain miscellaneous factors will give better prediction of occupational aptitude than will tests alone.

**In lieu of tests.** There are employment situations in which it is not feasible to embark on a scientific testing program with a view to developing employment technique. Perhaps the concern cannot afford, at the time, the necessary outlay or it is inadvisable to take the employees away from their work long enough to test them. Perhaps the present number of workers is too small for statistical purposes, but records of a biographical nature and production figures are available for a larger number of former employees. In such instances some of these miscellaneous factors may be used in lieu of tests and prove better than nothing. Moreover, there are various methods ordinarily in unsystematic use, such as letters of application, recommendations, and interviews, which can be systematized to advantage or can be evaluated statistically to determine whether they are actually worth using at all. The following factors will be discussed in the present chapter: academic record, initial success in vocation, personal history blank, letter of application, recommendations, and the interview.

#### ACADEMIC RECORD

It is often a simple matter for the employer to obtain a transcript of the applicant's academic record in school or other educational institution. Many application blanks call for the grade finished in school. But while this may give a rough indication of educational attainment, it is doubtless better to obtain school marks or something analogous. Where the situation warrants, it is often possible to write to the institution which the applicant at-

tended and obtain information regarding his educational career. This practice is especially common in the case of persons who have attended technical institutions and apply for positions along the technical lines pursued.

**School progress a selective procedure.** There are *a priori* grounds for believing that school progress should give some such indication of subsequent success. The school itself has probably exercised a certain amount of selection among its pupils. There are some individuals who are able to meet the normal educational demands and progress at the ordinary rate. Some, however, are unable to meet those demands and fall behind or perhaps drop out rather early in their educational career. Others, on the contrary, may be able to progress more rapidly because of their superior capacity. Thus indirectly the rate of progress in school, especially with reference to advancement or retardation, gives some indication of capacity to meet the problems and demands of the school situation. Similar principles apply to the actual grades or marks received in school. These should in the long run reflect the student's actual accomplishment and this in turn give some indication of his ability. These suggestions must be qualified in the light of the fact that students do not always use the ability which they possess and hence their grades may be an unreliable indication of that ability. Moreover, if a school system is poorly organized with inadequate methods of grading or promotion, little significance can be attached to the results. However, in the general case there is some ground for the assumption that the school curriculum is after all a rather prolonged mental test.

**Early academic record prognostic of later.** Various statistical studies have been made to determine the prognostic value of school marks. For instance, it has been shown that grades obtained early in the academic career are quite indicative of marks obtained later in the career. Much of this sort of data is available, but only two or three cases will be cited. (Cf. 246, 177.) Records were obtained of pupils in the 4th, 5th, 6th, and 7th grades and their average marks in each grade correlated with marks in the first year of high school. These correlations were as follows:

4th grade.....	.62
5th grade.....	.53
6th grade.....	.73
7th grade.....	.72

Evidently marks in the 6th and 7th grades give a pretty fair indication of marks in freshman year of high school. In another study of the relation between average marks in elementary school and average marks in high school the correlation was .71. Similar results are obtained when success in high school is compared with that in college. In one instance the group was divided into an upper and lower half on the basis of college marks and likewise on the basis of high school marks. It was found that about seventy per cent of those in the upper high school half were in the upper college half, while about the same per cent of those in the lower half in school were in the lower half in college. In another similar study the conclusion was drawn that three fourths of those entering the university from high school maintain approximately the same rank that they had in high school. Or, again, if a student is in the upper quarter in high school the chances are about four out of five that he will be in the upper half in the university.

The same principle seems to hold for predicting success in graduate work in professional schools from undergraduate college marks. A group of Harvard College graduates who also graduated

TABLE LII. PER CENT OF STUDENTS GRADUATING FROM COLLEGE WITH HONORS WHO RECEIVE DISTINCTION IN LAW AND MEDICAL SCHOOL<sup>1</sup>

COLLEGE DEGREE	CUM LAUDE IN LAW	CUM LAUDE IN MEDICINE
Plain.....	7	36
Cum laude.....	22	76
Magna cum laude.....	40	87
Summa cum laude.....	60	100

<sup>1</sup> After Lowell.

from law or medical schools was studied to determine the relation between taking honors in college and doing likewise in the professional school. (346.) Table LII indicates that of those who entered law school with a plain degree — i.e., with no distinction — only

DETERMINANTS OF VOCATIONAL APTITUDE 375

seven per cent received a degree with distinction in law and the corresponding figure for medicine is thirty-six per cent. However, of those who had a college degree "cum laude" — i.e., with distinction — twenty-two per cent received a degree with distinction in law, and of those who similarly entered medical school seventy-six per cent received the "cum laude." The graduates who received "magna cum laude" — i.e., high distinction — have an even better record in the professional school, and those with "summa cum laude" — i.e., highest distinction — have the best record of all in the professional school — sixty per cent of those in law taking honors and all of those in medicine doing so. The foregoing illustrations are sufficient to indicate the possibility of using early success in school as an indication of later success in academic lines.

**Academic record and occupational success.** The more important problem from the employment point of view is, of course, the extent to which school marks may be indicative of subsequent proficiency in industrial or professional activities. A study of graduates of Wesleyan University throws some light on this problem. (424.) The students who graduated between 1860 and 1889 were divided into three groups—those who graduated with valedictory or salutatory honors, i.e., ranked either first or second among their graduating classmates in scholarship; those who were elected to Phi Beta Kappa, an honorary fraternity for which high scholarship is the prerequisite, and the remainder who achieved no such

TABLE LIII. PER CENT OF COLLEGE GRADUATES FOUND IN "WHO'S WHO" <sup>1</sup>

GROUP	PER CENT
Honor men . . . . .	48
Phi Beta Kappa . . . . .	31
Others . . . . .	10

<sup>1</sup> After Nicolson.

honors. The per cent of each group appearing in the 1914 edition of *Who's Who* was then computed. These per cents are given in Table LIII. It is obvious that the honor men and the members of Phi Beta Kappa stand a much higher chance of distinction of the

sort under consideration. The group which took no academic honors or distinction constitutes about two thirds of the entire group, but actually contributed only about one third of the graduates who appear in *Who's Who*. To be sure, the type of success that lands one in *Who's Who* is apt to be literary, professional, political, or academic rather than industrial or commercial. Unfortunately, such a clear-cut criterion is not available for these latter types of success. There is some presumption, however, that if a criterion were available similar results would be obtained because intellectual factors are involved in all these types.

A study was made of the graduates of a technical institute in mechanical and electrical engineering, comparing marks at the institute with subsequent salary. Men of the graduating classes of three successive years were studied and their salaries obtained from four to six years after graduation. While success in engineering vocations may not be entirely reflected in salary, and other factors besides proficiency may influence salary, nevertheless, it gives some indication of vocational success. The men were divided into four groups on the basis of their school marks and the average salary obtained by each group was computed. The results are shown in Table LIV. (246, 198.) To facilitate comparison the salary of the

TABLE LIV. SCHOLARSHIP IN A TECHNICAL INSTITUTE AS INDICATIVE OF SUBSEQUENT SALARY <sup>1</sup>

SCHOLARSHIP GROUP	AVERAGE SALARY	PER CENT SALARY
Highest quarter.....	1664	100
Second quarter.....	1462	88
Third quarter.....	1418	85
Lowest quarter.....	1279	77

<sup>1</sup> From Hollingworth's *Vocational Psychology*, by permission of D. Appleton and Company, New York.

highest group is taken as 100 per cent and the others reduced to per cents thereof. It is obvious that the men who had better records while at the institute obtained appreciably higher salaries on the average. (These salaries were in 1913.) The differences are not large, perhaps, but enough to indicate the trend. If the data are handled by correlating individual salaries with individual marks,

## DETERMINANTS OF VOCATIONAL APTITUDE 377

the correlations for the graduates for each year are all positive and average .27. This indicates some relation between the two variables in question. The correlation is by no means sufficiently high to warrant academic record in such instances being used as the sole means of predicting vocational aptitude, but such record might prove of some value, as above suggested, in supplementing other indications.

The academic records of over 4000 graduates of West Point from 1818 to 1905 were studied with reference to subsequent success. (456.) The criterion of success was taken as appointment to the rank of Brigadier General or above. The members of each graduating class were divided into four groups of equal size on the basis of scholarship and for each group the number noted who achieved success in the above sense. The figures for all the first quarters were totaled, likewise the figures for all the second quarters, etc. It was then possible to compute for all graduates who had stood in the first quarter of their class what per cent were successful and to make similar computation for all the other quarters. The results are shown in the first part of Table LV. Whereas twenty-nine per

TABLE LV. SUCCESS OF WEST POINT GRADUATES IN RELATION TO SCHOLARSHIP <sup>1</sup>

CLASSIFICATION	PER CENT SUCCESSFUL
First quarter.....	29
Second quarter.....	24
Third quarter.....	21
Fourth quarter.....	15
First man.....	47
Second man.....	38
Next to last man.....	14
Last man.....	6

<sup>1</sup> After Covell.

cent of those in the highest quarter in scholarship achieved the rank of Brigadier General, only fifteen per cent of those in the lowest quarter did so. The results are more striking when we consider only the men at the extremes of their graduating classes in scholarship. From the lower part of Table LV we see that of all the men who ranked highest in their graduating class forty-seven

per cent were successful. The per cent was a little smaller for the men who ranked next to the top in their class. In contrast with these are the figures for the two lowest men in their graduating classes. The man who ranks at the top of his class stands almost eight times as great a chance of success as does a man who ranks at the bottom of his graduating class. Further evidence may be obtained by considering the men who were dismissed from the army. Of those so dismissed, eighteen per cent had been in the first quarter of their graduating classes, sixteen per cent in the second quarter, twenty-seven per cent in the third quarter, and thirty-nine per cent in the lowest quarter. Dismissal, of course, depends on various moral factors as well as ability, but there is an indication that these military failures were drawn largely from those with the poorer scholarship records. It is evident that success in this particular line could to some extent be predicted on the basis of scholarship.

**Amount of education.** It is common practice in obtaining information from an applicant to ask what grade in school he finished. In many instances it is not feasible to get the actual academic marks, but merely a statement of how far in the ordinary educational curriculum the individual progressed. This, however, has some significance. In many cases the employer is interested in whether the applicant has certain educational fundamentals which will be actually necessary for his work. He may need a certain amount of arithmetic, such as fractions, in order to make out time slips or compute dimensions of material that is to be used. He may need a certain proficiency in reading in order to interpret type-written directions or orders that are issued. If he has not progressed beyond a certain grade in school, it is probable that he has not been exposed to fractions or to reading of the requisite difficulty.

There is another aspect of the matter that is significant with the younger generation. In these days of compulsory education the grade finished in school is an indirect indication of intelligence. Suppose that in a given State every one is compelled to attend school until the age of sixteen. If, then, one individual has finished the third year of high school and another only the seventh grade, both having attended school some eleven years, it is obvious that the latter has occasionally failed to be promoted. This may indi-

cate poor teaching or improper motivation by parents and others, but it also probably indicates a difference in the innate intellectual capacity of the two persons. The same information may, of course, be obtained, if the applicant's statement can be trusted, by inquiring both the grade completed before leaving school and the age at leaving. The tendency for pupils of high intelligence to progress more rapidly in school when given opportunity has been repeatedly demonstrated. Rapid progress may then give some presumption of greater intellectual capacity. Hence these data as to the grade reached in school in a given time may afford an indirect approach to the same thing that is approached by the intelligence tests. In situations where tests are not used, some inkling as to the applicant's intelligence may be obtained in this fashion.

A few cases will be cited in which the amount of education was compared with an occupational criterion. There were indications in studies of aviation cadets and of aviators that the amount of education bore some relation to success at the aviation ground school or at the flying school. The correlation between amount of education as stated by the candidate himself and his average mark in the work of the ground school on engines, gunnery, signaling, theory of flight, etc., was .35. (624, 609.) With students of radio mechanics during the war it was found that "schooling is one of the best diagnostic criteria in selecting men to be trained in the care and repair of wireless apparatus." (608, 117.)

With a group of billing-machine operators the number of years of schooling gave a correlation with speed in billing of .23 and with accuracy in billing of .31. (308.) These correlations are small, but of some interest. With clerks in an insurance company the correlation of years of schooling with grade of work was .47. (612.) With students of telegraphy no correlation at all was found between years of schooling and receiving ability after 100 hours of practice. (608.)

An adding-machine company found that fifty per cent of its superior salesmen were college men, thirty per cent had attended high school or business school, while twenty per cent had only a grade school education. However, when all the men of the sales force were considered, only twelve per cent of the college men were



"A" salesmen, while twenty per cent of the grade school men were in this class. (278, 224.)

In another concern forty-five per cent of the successful salesmen were college men and only thirty-five per cent of the failures were college men. Another company found that men with high school education made more successful salesmen than those with more or less than this amount of education. This seemed true in some insurance companies, but in another group of insurance men there was a correlation of only .11 between years of schooling and production and the college men seemed best, grade school men the next best, and high school graduates the worst.

Findings such as the foregoing point to the necessity of evaluating a particular variable, such as education, with reference to the particular situation in which the variable is to be used. Such a factor may be of some value for vocational prognosis in one organization and worthless in another.

**Academic record in special subjects.** While general educational attainments give indirect evidence regarding intellectual capacity there is a further possibility that effort or achievement in special educational subjects may afford some indication of special capacity or interest that will be of vocational significance. High marks in a particular subject, such as English or mechanical drawing, may indicate special aptitude in that line. In elective courses the choice of certain subjects may indicate at least an interest and perhaps also some ability in those subjects.

The vocational implication of some of the more extreme cases is obvious. A person who has shown aptitude for mathematics by achieving good grades in his mathematics courses will qualify, other things being equal, for industrial work in which it is necessary for him to make computations. Conversely, a person who has failed in most of his mathematics, manifesting thereby an inaptitude for that kind of work, will perhaps be less effective in an occupation where considerable computation is involved. Similarly, a person who, according to school records, has done well in manual training has thereby demonstrated some mechanical proficiency and the expectation of his being successful in mechanical work is consequently somewhat greater. On the other hand, an individual with

obvious mechanical inaptitude as revealed in his academic record is apt to be a failure in an occupation that is exacting from a mechanical standpoint.

If a person has had the possibility of electing certain school subjects rather than others, and if it can be established that he made his own choice without undue influence by relatives or acquaintances, those choices will reflect either his ability or his interest or both. The average pupil selects school subjects which he likes and usually those in which he is fairly proficient. Persons who, for instance, have voluntarily chosen to study mathematics or some of the natural sciences will perhaps stand better chances in engineering occupations than will students who of their own choice pursued history or the classics. In this connection, however, it is essential to determine whether the choice was the applicant's own or whether it was the result of influence by other people. While it has been discovered that some of the most successful engineers are persons with a classical education, this merely reflects the fact that their parents had been of a high order of intelligence, had consequently obtained a liberal education in the day when only the more intellectual went to college, and had then encouraged their children to pursue the same type of classical education. These children inheriting the high intellectual capacities of their parents were destined for reasonable success in almost any line they might pursue.

To cite a practical instance of the use of academic record in special subjects, the application blank for candidates for aviation contained the direction, "Give names of the three studies in which you did the best work in the last two years of school." (624, 612.) The answers were evaluated as follows: a credit of +1 was given for each entry of physics, chemistry, or an engineering subject, and a penalty of -1 for each entry of Latin, English, history, philosophy, language, and the like. The resulting scores correlated with achievement in the aviation ground school to the extent of .28.

Inferiority of academic record to actual tests. These aspects of the academic record should not be used to the exclusion of the quantitative measurements which have been described in previous chapters, unless absolutely necessary. It can be shown that these

factors are not as valid as are mental measurements in predicting occupational success. Even if school records are available in quantitative form, so that it is unnecessary to take the applicant's word as to his educational career, it has been shown that these records are considerably less satisfactory than mental tests.

Data were obtained to indicate the relative validity of high school marks and of specific tests in predicting success during the first two years in an engineering college. (606.) Table LVI

TABLE LVI. CORRELATION OF FIRST TWO YEARS' WORK IN ENGINEERING COLLEGE WITH HIGH SCHOOL GRADES AND WITH SPECIAL TESTS <sup>1</sup>

HIGH SCHOOL GRADES	S.P.E.E. TEST SCORES
Algebra..... .21	Arithmetic..... .45
English..... .21	Algebra..... .30
Geometry..... .22	Geometry..... .35
Physics..... .25	Intelligence..... .29
Chemistry..... .29	Physics..... .36
Total..... .28	Technical information.. .22
	Total..... .48

<sup>1</sup> After Thurstone.

gives these correlations. The left part of the table shows the correlations between high school grades in algebra, English, geometry, physics, and chemistry with grades in the first two years of the engineering college. Algebra is the least predictive and chemistry the most. When the marks in these subjects are combined, the total correlates with the college work to the extent of only .28. These same students were given the test for engineering aptitude devised by the Society for the Promotion of Engineering Education. This test comprises six parts, each occupying about thirty minutes. The correlations of these appear in the right part of the table. It is to be noted that the thirty-minute test for arithmetic gives the best prediction of any single measure. It is also to be noted that in every instance a thirty-minute carefully standardized test dealing with specific information in a school subject is more predictive than the entire high school record in that particular subject. High school grades in algebra, for instance, correlate .21 with college grades, while the special algebra test correlates .30. The correlation of the total test score with college work is .48.

## DETERMINANTS OF VOCATIONAL APTITUDE 383

This may be contrasted with the corresponding correlation of .28 for the high school grades. This substantiates the point above mentioned that school grades are at best considerably inferior to actual scientific measures for vocational prediction. They should be used in lieu of them only when it is impossible to obtain the psychological measures. Whether or not they are valuable in supplementing such measures must be determined in the particular vocational situation.

### INITIAL AND SUBSEQUENT SUCCESS IN THE SAME OCCUPATION

In some instances it is possible, knowing a person's production record in a given occupation over a short period of time, to predict his subsequent efficiency. An investigation of this sort was conducted with insurance salesmen. (207.) In this case production figures were available for several groups of salesmen covering some length of time. The first group was small, but records were available for four years. Group II was larger and three years' records were available. Group III was larger still, but only two years' records were available. With these data it was possible to compare success the first year with success in subsequent years. The correlations between production in different years are shown in Table LVII.

TABLE LVII. CORRELATIONS BETWEEN SALES PRODUCTION IN DIFFERENT YEARS<sup>1</sup>

	GROUP I	GROUP II	GROUP III
First year and first subsequent year . . . . .	.92	.72	.74
First year and second subsequent year . . . . .	.76	.68	
First year and third subsequent year . . . . .	.47		
First year and total production . . . . .	.90	.88	.93

<sup>1</sup> After Goldsmith.

It can be seen, for instance, that with Group I the correlation between first-year production and production the subsequent year is .92, whereas the correlation between the first year and the

second subsequent year is .76 and that between the first year and the third subsequent year is .47. This same tendency is indicated in Group II; namely, the first year gives a better indication of the first subsequent year than of the later years. The figures at the bottom of the table which indicate the correlation of the first year and total production of all the years are quite large.

These same data may be presented from a different standpoint by merely dividing each group into two classes, those above the average and those below the average. It is then possible to note what per cent of those above the average the first year remain subsequently in that same superior class. These figures are as follows:

- Group I.** Full-time men whose records are complete for four years:
- 100 per cent of those beginning above the average remain above the average in one of the three succeeding years.
  - 100 per cent of those beginning above the average remain above the average in two of the three succeeding years.
  - 80 per cent of those beginning above the average remain above the average in all of the three succeeding years.
  - 100 per cent of those beginning above the average have a total production above the average for the four years.
  - 93 per cent of those beginning below the average remain below the average in one of the three succeeding years.
  - 93 per cent of those beginning below the average remain below the average in two of the three succeeding years.
  - 80 per cent of those beginning below the average remain below the average in all of the three succeeding years.
  - 93 per cent of those beginning below the average have a total production below the average for the four years.
- Group II.** Those whose production records are complete for three years:
- 89 per cent of those beginning above the average remain above the average during one of the two succeeding years.
  - 80 per cent of those beginning above the average remain above the average during both the succeeding years.
  - 91 per cent of those beginning above the average have a total production above the average for the three years.
  - 98 per cent of those beginning below the average remain below the average for one of the two succeeding years.
  - 83 per cent of those beginning below the average remain below the average for both succeeding years.
  - 93 per cent of those beginning below the average have a total production below the average for the three years.

## DETERMINANTS OF VOCATIONAL APTITUDE 385

In this particular study there is evidently a fairly close relation between the initial and the subsequent production. If a man's early selling record is poor, some doubt may validly be raised as to the advisability of his continuing.

The number of accidents is sometimes taken as an inverse indication of a worker's efficiency. In a machine shop the correlation of the number of accidents in successive quarters of the year was computed; i.e., the tendency for a worker to have in a given quarter the same number of accidents which he had in the preceding quarter. Four such correlations for successive quarters are .72, .37, .53, .69. (322, 218.) A worker with a record of accidents is more liable to have others than is a worker with a clear record. One accident does not apparently make a person much more careful. The accidents seem to be due to some fundamental cause. In so far as they are an index of inefficiency, early failings in this respect are prognostic of later.

With billing-machine operators efficiency both from the standpoint of speed and accuracy in the sixth month of work was studied to see how well it could be predicted from earlier efficiency and also how well it would predict later efficiency. The correlations are shown in Table LVIII. The work during the first month

**TABLE LVIII. CORRELATION OF EFFICIENCY AT BILLING-MACHINE  
DURING SIXTH MONTH OF SERVICE WITH EFFICIENCY  
DURING OTHER MONTHS<sup>1</sup>**

	SPEED	ACCURACY
First month.....	.20	.00
Second month.....	.48	.25
Third month.....	.60	.31
Fourth month.....	.76	.58
Fifth month.....	.76	.74
Seventh month.....	.86	.72
Eighth month.....	.74	.40
Ninth month.....	.68	.16

<sup>1</sup> After Kornhauser.

is worthless from a predictive standpoint. From then on it appears somewhat diagnostic. This is especially the case with speed,

which becomes of some significance in the second or third month. Accuracy has little predictive value until almost the fourth month. The correlations of the sixth month with adjacent months are of course higher than with more distant months. (308.)

The foregoing are some typical studies of the relation between initial and subsequent success. Where a concern finds it impossible to engage in any systematic employment program, it may be worth while to keep production records of workers from the outset and determine whether there is a relation such as that existing in some of the cases just recounted. If such a relation is found to be rather close, workers with poor initial records may well be considered for transfer to some other line for which they are better adapted.

#### PERSONAL HISTORY OR APPLICATION BLANK

A personal history or application blank is often filled out as a preliminary to an interview. This is sometimes desired in order to save the interviewer's time, and sometimes to sort in a preliminary way from a group of applicants those who are worth interviewing at all. The blank aims to bring out the more obvious data regarding a person's capacities and interests and may form a basis for the subsequent securing of more detailed information. It may be filled out entirely by the applicant or else partially or entirely by the interviewer. It is sometimes arranged so that the applicant fills one side while the interviewer uses the reverse. At any rate, some form of such a blank is found in most employment offices.

**Technique of evaluating items in blank.** Such personal history blanks are generally used rather uncritically. It is assumed, perhaps on the basis of casual observation, that certain items, such as age or marital status, are prognostic of occupational success. This assumption may not be as erroneous as the assumption that a bump on the head just above the ears indicates ability at constructing things or that fine-textured skin presages artistic achievement. But it is, nevertheless, an assumption, whereas science prefers to deal with facts. It is possible to evaluate these various items of personal history statistically and get the facts. After a group of individuals have been on the job sufficiently long to demonstrate their ability, it is possible to determine whether certain items

actually differentiate the good from the poor workers. If, for instance, a group of salesmen are divided into successes and failures, and it proves that most of the successful salesmen are married and most of the failures are single, this information as to marital status may be of some significance in employing salesmen in the future. Of course, it is not feasible in this type of problem to employ any very rigorous statistical analysis. About the best that can be done is to divide the individuals into two classes as far as the criterion is concerned. These two classes may simply represent a division at the mid-point of the range of occupational ability, or it may consist of classes at the two extremes. Having made such a division, however, it is a simple matter to tabulate for any particular item on the history blank the per cent of the successful group giving a certain answer or failing to give it and similar percentages for the unsuccessful group. If a particular answer is given much more frequently by the successful individuals than by the unsuccessful, that item or answer may be taken as to some extent differential of success in the occupation in question. If there is any doubt as to whether the difference is large enough to be significant, recourse may be had to proper formulæ for determining this significance. (Cf. p. 307.)

**Physical characteristics.** The usual application blank calls for such items of a physical character as height and weight. It is well then to determine whether there is any relation between these and fitness for the particular job. Of course there are certain instances where it is perfectly obvious that large stature is desirable. If a man has to lift the iron core to a considerable distance from the floor in order to place it on the tire-building machine, height is an obvious requisite unless hoists are provided. In hauling a heavy truck or doing work where great force must be exerted, a large man has an obvious advantage. Such patent cases as this, of course, need no scientific study. There are, however, more subtle possibilities in stature. One that has sometimes been rather seriously considered by employers of salesmen is the possibility that large salesmen can "dominate" the prospect and hence make more sales. One manager actually attempted to develop a sales personnel over six feet in height. Some of us have occasionally felt



ourselves shrink and weaken in the presence of a large man who was leading us up to the dotted line. If we are standing and it is necessary to look up at him in the conversation, this mere bodily posture is liable to have psychological effects and the upturned chin with its subtle suggestion of inferiority renders one more vulnerable to a verbal uppercut. The writer makes it a practice when interviewed by a salesman in the 200-pound 6-foot class to have the man seated, and if possible to sit on the desk himself, so as to dominate the salesman rather than to be dominated by him.

Some statistical evidence is available on this matter of stature. In two concerns salesmen were divided into approximately three classes of equal size on the basis of their sales records. (285.) The average height and weight of each group is given in Table LIX.

TABLE LIX. AVERAGE HEIGHT AND WEIGHT OF SALESMEN OF DIFFERENT DEGREES OF EFFICIENCY <sup>1</sup>

SALES RECORD	COMPANY A		COMPANY B	
	Average height (inches)	Average weight (pounds)	Average height (inches)	Average weight (pounds)
Highest third.....	69.0	156	69.3	180
Middle third.....	68.6	153	68.0	185
Lowest third.....	69.8	158	69.0	178

<sup>1</sup> After Kitson.

Within these groups there is evidently little relation between stature and selling. In height the group of medium selling ability has the lowest average. The weights likewise are equivocal. In one company the poorest salesmen are slightly the heaviest, while in another the medium group seems most ponderous.

Results of this sort, however, are not always found. (278, 219 ff.) An insurance company found that the average monthly sales of men under 69 inches in height were \$740, while the sales of those over 69 inches were \$1165. In another concern the average height of the ten leading salesmen was 70.7 inches, while the average height of all failures was 69 inches. It was found in another group that

## DETERMINANTS OF VOCATIONAL APTITUDE 389

men weighing between 140 and 180 pounds averaged higher in monthly production than those above or below those limits.

This is not the whole story, however. While there may be no universal tendency within a given sales group for the larger men to be the more effective, the evidence is clearer that salesmen as a whole are larger than the average individual. The results of a number of studies, including the one given in the preceding table, are summarized in Table LX. The average height and weight are

**TABLE LX. AVERAGE <sup>1</sup> HEIGHT AND WEIGHT OF DIFFERENT GROUPS OF SALESMEN <sup>2</sup>**

	AVERAGE HEIGHT (inches)	AVERAGE WEIGHT (pounds)
Mixed group.....	69.6	170
Insurance.....	69.5	
House-to-house.....	69.5	158
Technical.....	69.3	169
Miscellaneous A.....	69.1	155
Miscellaneous B.....	68.8	181
Miscellaneous C.....	69.5	160
General population — army....	67.5	142
General population — actuarial data.....	68.5	

<sup>1</sup> For the smaller groups the median was used rather than the ordinary average. It is found by arranging the measures in order of magnitude and then selecting the middle one in the scale. In small distributions it prevents an extreme case from seriously affecting the average.

<sup>2</sup> After Kenagy and Yoakum, *et al.*

given for various sales groups. For comparison with the general population the average height and weight of about 1,000,000 men in the army is included. The average height of some 220,000 men tabulated by the Association of Life Insurance Medical Directors is also given. This is somewhat larger than the army average, but not as large as the average of any of the sales groups. These latter are considerably superior to the general population in both height and weight.

The difference should be somewhat qualified in the light of the fact that the army group was somewhat younger than the others. Most of the groups of salesmen average in their thirties, while the average man in the draft was well below this. Many persons, of course, put on weight as they grow older and the salesmen might

have been heavier in part because of their maturity. It is rather doubtful if this would account for differences of twenty pounds or more in the average case. Moreover, results for height would be much less affected by this error because this characteristic changes very little after reaching maturity.

So, while there may be some doubt regarding the relation of stature to production within a given sales organization, there is no question but that salesmen as a whole are larger than their prospects. If there is any conclusion to be drawn other than this, it is that perhaps men of medium stature, although above that of the general population, are somewhat more effective than those at either extreme. It has been more or less seriously suggested that such a salesman is large enough to dominate his prospect effectively, but not too large to get around easily and cover the ground.

**Age.** Considerable significance is attached to age in employment and analogous problems. Some railroads will not employ a man who is over 35 and retire employees on a pension at the age of 65 or 70. Similar retirement is sometimes applied to the teaching profession. Some States will not permit a person under 16 to drive an automobile. A person must be 21 in order to vote. In some States minimum age limits of 14 to 16 are set, below which an individual cannot be employed in industry.

Such tendencies are based usually on popular belief that persons outside of the age limits in question are ineffective in the type of work under consideration. It is in point then to consider more systematically any psychological aspects of age that may be of vocational significance. We know, of course, that mental proficiency does change in one's early years and the changes at the other extreme are obvious. The influence of age on performance in certain mental tests was mentioned in Chapter VII. (Cf. Figure 3, p. 180.) Proficiency in all the tests increased from childhood up into the teens. However, the rate of increase was not uniform. For instance, the type of coördination shown in a tapping test progressed more rapidly than sheer muscular strength with the implication that persons in their early teens are better suited to work requiring rapid coördination than to work requiring muscular strength. Likewise at the other extreme as far as the tests were

## DETERMINANTS OF VOCATIONAL APTITUDE 391

applied there appeared a decrease in proficiency in middle life at the type of performance measured by a test of substituting symbols for numbers. It is quite possible that rather extensive age differences of this sort exist and if so some of them may be of vocational significance.

An obvious approach to the problem from the practical standpoint is to correlate age with occupational proficiency and to determine within a particular group of employees if the more mature are the more proficient. With clerical workers in the civil service a correlation of .06 was found between age and efficiency scores. (166.) With another group of clerks a correlation of .35 was found between grade of work done and age. (612.) With a group of telegraphers the correlation between age and receiving ability was -.09. With insurance salesmen production correlated with age at the time of initial contract with the company to the extent of .15. Only one of these coefficients is large enough to be of any possible value. This does not give the whole story, however, because it may be that persons of medium age are most efficient rather than the oldest ones, whereas a large correlation would not be obtained unless the oldest ones were the best.

This factor may be investigated by noting the relative efficiency of workers of different ages, with a view to determining whether for a given occupation there is an optimum age. With a miscellaneous group of superior salesmen the average age was almost 39. Only 11 per cent of them were under 30 and only 10 per cent over 50. (278, 217.) Those of middle age were manifestly the big producers. This, of course, suggests that the younger men had not had sufficient experience and this is to some extent the case. In an insurance company where men with previous insurance experience were generally more efficient and where the best producers were between 35 and 50, it developed, nevertheless, that the best producers at the time of contract — many of them without previous experience — were between the ages of 30 and 45. Even apart from experience it seemed that maturity was desirable. Similar studies with other groups of salesmen have revealed the fact that extremes of age are somewhat less favorable than the middle range.

It might often be worth while with other kinds of occupations to

apply similar technique and determine whether there seemed to be any optimal age at the time of initial employment. There are doubtless many types of work in which maturity is necessary in order to impress favorably persons with whom one deals and there are other types in which a man who is too old will fail because of decreased mental efficiency. It is necessary to answer the question statistically in any given case.

There is another aspect of age that should be mentioned. Quite apart from efficiency there is a possibility that age may bear some relation to stability or turnover. Various studies have been made of this relation, but rather than revealing any specific effect of age as such they have brought out various other complicating factors that enter into different age groups. For instance, in a paper manufacturing concern there seemed to be considerable instability with the very young female employees and this was not remedied by any wage adjustment such as a bonus system. The fact was brought out, however, that in many instances the girls took their pay envelopes home unopened and their mothers received the bonus. There had been little motivation even from the start.

A study was made of workmen who quit in two large firms, one doing metal work and the other manufacturing furniture. (284.) These "quits" were classified as to age in five-year intervals and records tabulated to show the average number of weeks worked by individuals in a given group before they left the employ. Table LXI gives the results. The employees of company B are manifestly less permanent for all age groups. The important point in the present connection is the relative permanency of employees of different ages within a company. There is in both instances a manifest turnover among the younger workers. This doubtless reflects the natural instability of youth and the legitimate search for a vocational objective. At the other extreme there is a marked stability for those older than fifty. At this time one's interests have become fairly well established and profitable change in employment is rather unlikely. Likewise, between thirty and thirty-five there seems to be considerable stability, this being a period when many individuals buy homes or raise families. From then on until fifty there is somewhat of a decrease. It is quite possible

## DETERMINANTS OF VOCATIONAL APTITUDE 393

TABLE LXI. AVERAGE NUMBER OF WEEKS WORKED BY EMPLOYEES WHO QUIT<sup>1</sup>

AGE	AVERAGE NUMBER OF WEEKS	
	COMPANY A	COMPANY B
Under 21.....	18	10
21 to 25.....	19	9
26 to 30.....	23	11
31 to 35.....	31	24
36 to 40.....	28	19
41 to 45.....	29	12
46 to 50.....	30	8
51 to 55.....	58	15
Over 55.....	56	25

<sup>1</sup> After Kitson.

that at this period the worker's family is becoming more self-supporting and his domestic responsibilities are not quite so pressing. He then realizes that old age is coming soon and that he had better change his occupation now if at all. Consequently, he takes this opportunity to try other occupations with a view to a location which will be permanent and satisfactory. While this is only a single study, it is quite probable that these same mental factors of youthful instability, domestic responsibility, and subsequent search for ultimate occupational status are operating in a good many concerns and that in hiring employees with a view to turnover these facts may be of significance. They point incidentally to the desirability of watching for symptoms of unrest at these critical ages, being more tolerant of the workman and attempting to make such adjustments as will keep him on the job if he is actually satisfactory.

**Marital status.** Many employment men make it a practice to hire if possible married applicants and preferably those with addi-

tional dependents. The assumption is that such persons, because of their greater economic necessity, have greater incentive to do satisfactory work in order to hold the job and advance.

Practically the only available statistical studies of this factor involve salesmen. A number of such investigations are summarized in Table LXII. The preponderance of married men

TABLE LXII. PER CENT OF SUPERIOR SALESMEN THAT ARE MARRIED AND SINGLE <sup>1</sup>

GROUP	MARRIED	SINGLE
Miscellaneous.....	93	7
Insurance.....	94	6
Routine.....	61	39
House-to-house.....	81	19
Technical products.....	91	9

<sup>1</sup> From Kenagy and Yoakum's *The Selection and Training of Salesmen*, by permission of The McGraw-Hill Book Company, Inc., New York.

among the superior salesmen is obvious. It is particularly so with the higher types of selling. (278, 225.) In another instance in a single company 74 per cent of the successful salesmen were married and only 57 per cent of the unsuccessful salesmen. With a group of insurance salesmen who had not been with the company over two years, so that the factor of experience did not enter appreciably, the ratio of the average sales of the single group to that of the married group was \$9386 to \$10,000. (573.)

The results seem to indicate a rather marked superiority of married men as salesmen. One source of error in the generalization should be noted, however. The married men as a rule are older. Census figures indicate that of a random selection of white men between 25 and 29 years of age approximately 57 per cent are married, while of a similar group between 30 and 34 years of age about 75 per cent are married. (Cf. 322, 227.) We saw previously that the best salesmen were over 30. Hence the present results may to some extent be due to the fact that the older men prove more

efficient and also get married. However, three of the groups of superior salesmen listed in Table LXII show an incidence of marriage well over the census figures for other men in their early thirties. With the routine and house-to-house groups the salaries are almost too low for the support of a family and we find more single men. These groups are also recruited frequently from college students. Even here, however, the married men are the larger producers. As evidence from a slightly different angle we may mention the fact that with married insurance salesmen those whose wives were engaged in a gainful occupation produced only 70 per cent as much as those whose wives were dependent. The question might, of course, be raised as to which was cause and which was effect. Another insurance concern which found results like the foregoing, but found that marital condition at the time of contract was not differential, discovered, however, that the greatest improvement in selling was made by the men who were single at the time of contract, but had married since joining the company. This would indicate rather clearly the family incentive.

**Dependents.** If being married serves as an incentive for occupational effort one would expect other dependents to provide a similar motive. The same groups of superior salesmen recorded in the preceding table averaged about 2.5 dependents. (278, 226.) This is more than a wife, but it is not a large family. In another company the average number of the dependents of the successful salesmen was 1.9 and of the unsuccessful 0.8. Among a group of insurance salesmen those who were married but childless were slightly inferior in production to those who were single. However, the production of those with 1 or 2 children, with 3 or 4 children, and with 5 or more children were in the proportions, \$10,000: \$8792: \$7584. The man with children, but only one or two of them, seemed superior.

**Previous experience.** It is common practice to ask an applicant regarding his previous vocational history either in general or in work similar to that proposed. Of course, if the past work has been identical with the proposed — for example, wood heeling — the case is clear. The amount of experience the applicant has had in that type of work will be somewhat indicative of his proficiency.



In work at a trade it is, of course, desirable to develop a trade test (*infra*) instead of relying on the applicant's statement of ability or inferring proficiency from mere length of service. But even where a trade test is not possible, the amount of previous experience may be significant. As shown above, production in selling insurance the first year was somewhat predictive of selling in subsequent years. It is not safe, however, to assume that any previous kind of selling qualifies one for a particular sales job. From the considerations in Chapter X certain types of selling apparently require a person of higher intelligence than do others. A concern found to its surprise that it was getting its worst salesmen from applicants who had had more than five years' selling experience in other lines. (278, 235.) It was possible that the more experienced applicants had ultimately proved unsuccessful in the other lines and then applied to this concern. Another insurance company found that its best applicants had held some other, not necessarily a selling, position for several years, but had not remained so long with a former employer as to lose their adaptability.

The kind of job previously held may give some indication of success in a different proposed line. A concern studied carefully the previous occupations of its sales force with reference to their relation to turnover, length of service, per cent of dealers sold, and per cent of quota sold. Taking all these factors into consideration, the previous occupations were arranged into seven classes on the basis of the value of the class in predicting success in selling. The order of these classes was as follows: (1) professions, (2) business for self, (3) retail selling, (4) outside selling, (5) clerical, (6) minor executive, (7) trades. (278, 161.) Men recruited from the professions, however, had rather short length of service so that they constituted a rather unprofitable source of supply even though they were effective while they stayed. The next four in order constituted on the whole the best prospective material, while minor executives and tradesmen seemed a rather unprofitable source from which to recruit for this particular selling job.

**Miscellaneous factors.** Certain miscellaneous items of personal history may be of significance in a particular situation. For instance, with insurance salesmen the number of clubs to which a

man belonged was somewhat indicative of production. The correlation coefficient is small, but the amount of paid business solicited increases gradually with increasing number of clubs, and the men belonging to seven clubs had the best record of all. (360.) With insurance salesmen those who carried a considerable amount of insurance themselves proved more effective. Reasons for entering a vocation may be of some significance. It was found that employees who entered an occupation because of the influence of a friend were not as effective as those who had entered for ulterior reasons. Possibly this latter reflected a real interest in the work, while the former indicated mere accident.

**Combinations of personal history factors.** Where a concern has studied factors such as the foregoing, it is common practice to state the aspects of these different factors that are of significance. This is somewhat analogous to stating critical scores only in less systematic form. The important items are usually given without indicating their relative importance or attempting to combine them into a single figure. Two such descriptions for salesmen will be quoted by way of illustration:

In age he will range between twenty-six and thirty years at entrance; he will be tall, but his success wanes as he rises above six feet; he should be well built and certainly not so much as ten per cent underweight; the man of foreign birth is more successful than one of native birth and the American born man of foreign parentage is more successful than the American born of native parentage. A good education is a positive asset provided that education does not exceed the degree of A.B. or B.S. Previous selling experience is advantageous, but men who have sold for over five years are not so promising as those with a somewhat briefer experience. Membership in a fraternal order is an advantage if the member attends regularly, but no added advantage is evident from membership in a number of such affiliations. (573.)

The items of the personal history record related to success in selling were: selling experience other than insurance, insurance experience of one year or more, high school education, three or more dependents, election to office twice or more in social organizations, attendance at social affairs oftener than an average of once per week, participation in athletics of four or more kinds when in school, participation in other school activities and an ambition to attain an executive position in the insurance business. The items showing a significant relationship to lack of success in selling were: age of twenty-two years or younger, schooling of eight grades or less, membership

in two or fewer social organizations, no election to office in such organizations, and a purpose of continuing in the insurance business less than ten years. (468, 57.)

It is more scientific not merely to state these qualifications, but to indicate their relative importance. The best procedure is to weight the items in some fashion if possible. In a few cases efforts have been made to do this and to combine the factors into a total score. To cite the case of an insurance company, after the items of the application blank had been studied in the manner just described, the arbitrary weighting scheme shown in Table LXIII was devised. (207.) The weighting takes account of the fact that very young persons are not as apt to be successful in selling as are those

TABLE LXIII. WEIGHTS OF ITEMS OF PERSONAL HISTORY FOR PREDICTING EFFICIENCY IN SALESMANSHIP<sup>1</sup>

	WEIGHT		WEIGHT
<b>Age:</b>		<b>Education:</b>	
18 to 20	-2	8 years	+1
21 to 22	-1	10 years	+2
23 to 24	0	12 years	+3
25 to 27	+1	16 years	+2
28 to 29	+2		
30 to 40	+3	<b>Occupation:</b>	
41 to 50	+1	Social	+1
51 to 60	+0	Non-social	-1
Over 60	-1		
		<b>Own insurance:</b>	
<b>Marital status:</b>		Carried	+1
Married	+1	Not carried	-1
Single	-1		
		<b>Contract.</b>	
<b>Clubs:</b>		Full time	+2
Belongs	+1	Part time	-2
Not belong	-1		
<b>Experience:</b>			
Previous life insurance experience			+1
<b>Confidence:</b>			
Replies to question: "What amount of insurance are you confident of placing each month?"			+1
Does not reply to this question			-1

<sup>1</sup> After Goldsmith.

## DETERMINANTS OF VOCATIONAL APTITUDE 399

who are somewhat older. Hence those particular ages receive a negative weight and count somewhat against the applicant. On the other hand, an age between thirty and forty receives the most credit, whereas persons older than forty appear to be somewhat less valuable to the company. Similarly, with education there appears to be a maximum value for persons who devote about twelve years to their educational career. Married applicants receive more consideration than unmarried. Previous occupation seems significant when considered from the standpoint of whether or not the occupation involved social contacts, such as selling, work at a cashier's window, or reporting. It also appears that individuals who are contemplating full-time service are a better investment than those who are proposing to work only part time. Carrying insurance one's self likewise seems in the applicant's favor as does also belonging to various clubs. Other items to be noted are previous experience with life insurance and the applicant's own confidence as to what he will be able to do in actual selling.

When a large group of insurance salesmen were classified, on the basis of their production records, into a best group, a middle group, and a poorest group, and their scores on the personal history blank computed according to this weighting, the results are as shown in Table LXIV. The entries in the table are the per cent of individ-

**TABLE LXIV. PER CENT OF SALESMEN WITH DIFFERENT PRODUCTION RECORDS MAKING VARIOUS SCORES ON WEIGHTED PERSONAL HISTORY ITEMS <sup>1</sup>**

PRODUCTION RECORD	SCORE ON PERSONAL HISTORY BLANK		
	Below 4	4 to 8	Above 8
Best group.....	15	41	44
Middle group.....	18	54	28
Poorest group.....	53	37	10

<sup>1</sup> After Goldsmith.

uals in the given production group falling in the various classes of weighted score. The largest per cent of the best group score

above 8 points, while much smaller proportions of the two other groups do so. The poorest group has a majority of its scores below 4. On the basis of these results a critical score of 4 points was recommended. If applicants below this score were rejected, it will be seen that many of the inferior salesmen would be avoided and comparatively few of the better ones eliminated.

It is sometimes possible to apply correlation procedure to items of personal history and weight them according to a regression equation just as was done with tests of special capacity. This is, of course, feasible only when the factors involved are such as yield a considerable range of possible values. A correlation based on a variable that involves only two classes, such as married *vs.* single, does not fit into the theory of the regression equation. Such an equation for insurance salesmen proved to be:

$$X_1 = 3.2 X_2 + 9317 X_3 + 106 X_4 + 5534 X_5 + 26880$$

where  $X_1$  is production,  $X_2$  the amount of insurance carried at the time of contract,  $X_3$  the number of clubs to which the man belongs,  $X_4$  the age at the time of contract, and  $X_5$  the number of dependents at the time of contract. (360.) When the items are weighted according to this equation the coefficient of multiple correlation — i.e., the correlation of the weighted sum of these items with the criterion — proved to be .40. This is a considerably better prediction than could be made with any single item.

Such a correlation is not, of course, sufficiently high to justify its use as the sole basis for selection or in lieu even of the various tests that might have been developed. However, such a weighted personal history record might form a valuable supplement to any other predictive measures that were available. If the data are in such form that correlation coefficients can be computed and a regression worked out, the effort may prove worth while. In some instances it may be possible to include some of these personal history items in a regression equation along with tests.

The foregoing are some of the items of personal history that are available in the average application blank and that have been shown in some situations to be indicative of occupational success. Just as with many other predictive measures one cannot assume

that what proved true in one situation will do likewise in a different one. It is necessary to evaluate the items with reference to the special problem for which they are to be used. However, there are manifestly many items of personal history that will be of value in certain employment problems once their significance has been ascertained.

#### LETTERS OF APPLICATION

The first step in many employment situations is the solicitation of letters of application. Help-wanted advertisements often require this form of reply. Such letters serve a purpose similar to that of the application blank in making a preliminary sorting of applicants with a view to finding those in whose case further interview or investigation is desirable. If a grossly misspelled letter is received from an applicant for a stenographic position the matter ends right there. It is also possible in this way to get a line on individuals who are at a considerable distance and who do not care to come and apply personally unless there is a fair prospect of their being hired.

A letter of application differs from the usual application blank in that it insures less in the way of specific information. Instead of asking the applicant specifically for biographical data he simply writes the facts which he considers most pertinent in qualifying him for the position in question. This factor of giving the applicant a chance to express himself freely while omitting some significant items makes it possible, on the other hand, in the opinion of some employment men, to judge something regarding such things as neatness, ability to express one's self, or tendency to be systematic which might not be manifest in the answers to specific questions.

In dealing with such letters of application it is, of course, necessary to make some qualitative evaluation of them. The average employment man deals with such letters by the usual procedure of general impression. One letter may, of course, be markedly neater than another, thus giving the presumption that the writer of the former is the neater individual. However, the question arises as to the reliability and validity of such estimates of the individual from the letter. The reliability of the estimates involves the extent to

which a given judge agrees with himself if he makes the estimate on different occasions or the extent to which he agrees with other judges. The validity of the estimates denotes the extent to which the results correlate with some further criterion, such as production or judgment of persons who actually are acquainted with the applicant and are not judging him merely by his letter. These problems of reliability and validity of estimates suggest the earlier discussion of evaluation of estimates of mental traits from physiognomy as manifested in photographs. (Cf. Chapter II.)

Experiments on the reliability of estimates of this sort have been conducted. An advertisement for a bookkeeper and office assistant was inserted in a New York paper and from the letters of application twenty-five were selected for study. (244, 10.) The signatures were removed and the letters marked with a key symbol in view of the possibility that some of the persons estimating the letters might be acquainted with an applicant. These letters were then submitted to fifty judges — business and professional men and women, students, and clerical workers. These judges ranked the letters in order from best to worst; i.e., numbered them from 1 to 25, with reference to (1) intelligence, (2) reliability, (3) tact, and (4) neatness. The ratings on all of these four traits were made separately. In addition, ten of the judges repeated this same procedure a month later without reference to their earlier estimates.

A detailed presentation of the results is not worth while in the present connection. Suffice it that there was a rather marked disagreement among the judges. For simplicity's sake we shall consider only ten typical judges. The first letter, when estimated as to the intelligence of the writer, is located all the way from 3 to 17, where a rating of 1 indicates the most intelligence and 25 indicates the least. This same letter gets a rating of from 2 to 23 for tact; from 4 to 19 for reliability, and from 1 to 13 for neatness. In other words, some judges rate the applicant as inferred from his letter very much more highly in some of these traits than do other judges. Taking the next letter in the same fashion the ranks assigned by these ten judges vary for intelligence from 4 to 45, for tact from 1 to 24, for reliability from 4 to 20, and for neatness from 4 to 25. A third letter similarly gives estimates of intelligence from

5 to 25, of tact from 2 to 25, of reliability from 2 to 25, and of neatness from 3 to 18. These are sufficient to illustrate the disagreement between the different judges and are by no means atypical of the results of the other forty judges. If they had shut their eyes while considering the tact of the writers, the judges would have agreed with one another regarding the order of the letters about as well as they actually did with their eyes open. The same thing was true regarding the estimates of the writers' reliability. The situation is slightly improved with reference to neatness and intelligence, but not to any great extent. Thus it seems that the reliability of the estimates from the standpoint of the agreement of the judges with one another is rather low.

To study reliability from the other standpoint of agreement of the judge with himself, we may consider the results of the ten judges who repeated the ranking procedure a second time one month subsequently and correlate their two sets of estimates. Such correlations will show, for instance, whether a given judge ranks the same letter high in intelligence in both first and second trials and rates another letter low in intelligence in both cases. These correlation coefficients are given in Table LXV. This table shows, for instance, that with Judge A his initial arrangement of letters from the standpoint of intelligence correlates .59 with his arrangement a month later. However, his initial and final arrangements for tact correlate to the extent of only .40, while the corresponding coefficients for neatness and reliability are .50 and .67 and the average of these four correlations is .54. This average gives a fair notion of the reliability or consistency of Judge A. It represents a fair agreement between his two ratings. In other words, he has not utterly changed his mind. A glance at the figures in the table shows that Judges B, D, and I are rather effective from this standpoint, whereas C and F are distinctly inferior. These latter are manifestly inconsistent with themselves and apparently change their minds very appreciably. If they were hiring employees on the basis of letters of application, many a person's destiny would hinge on the day when his letter happened to arrive or on what the employment man had eaten for lunch.

The results may also be considered from the standpoint of the



TABLE LXV. CORRELATION BETWEEN ESTIMATES OF TRAITS FROM LETTERS OF APPLICATION MADE ONE MONTH APART<sup>1</sup>

JUDGE	INTELLIGENCE	TACT	RELIABILITY	NEATNESS	AVERAGE
A	.59	.40	.50	.67	.54
B	.72	.72	.73	.72	.72
C	.08	.40	.27	.38	.28
D	.72	.44	.65	.88	.67
E	.60	.63	.20	.44	.47
F	.31	.18	.23	-.14	.21
G	.44	.52	.46	.92	.60
H	.62	.31	.45	.51	.47
I	.65	.71	.73	.91	.75
J	.63	.42	.52	.71	.57
Average	.54	.47	.47	.60	.53

<sup>1</sup> From Hollingworth's *Judging Human Character*, by permission of D. Appleton and Company, New York.

agreement of the individual judges with the consensus of opinion of all the judges. Some apparently agree more closely with the consensus than do others, but it proves rather difficult to locate an expert, i.e., one whose individual opinion is tantamount to the combined opinion of all. The results of this study then are a trifle discouraging from the standpoint of the consistency of the estimates of the judges based on letters of application.

It is rather probable that the unsatisfactory character of the results is partially due to the halo effect mentioned in the previous chapter. As a matter of fact in the present study it was found that there was a high intercorrelation between the traits. These intercorrelations are given in Table LXVI. Each entry represents the correlation between the trait listed at the end of the row and at the top of the column in which that entry appears. The correlations of intelligence with tact and reliability and of tact with reliability are

DETERMINANTS OF VOCATIONAL APTITUDE 405

TABLE LXVI. INTERCORRELATIONS OF ESTIMATES OF TRAITS BASED ON LETTERS OF APPLICATION<sup>1</sup>

	TACT	RELIABILITY	NEATNESS
Intelligence.....	.92	.91	.82
Tact.....		.91	.85
Reliability.....			.84

<sup>1</sup> After Walton.

over .90 and the other correlations are over .80. Manifestly the rater found it difficult to discriminate intelligence from tact or from reliability or to discriminate any of the traits from one another when making his estimates on the basis of the letter of application. It was largely a matter of general impression.

Experiments on the validity of estimates based on letters of application have been conducted in one instance. (458.) Twenty-five seniors in a school for religious workers wrote personal letters of application for positions of the sort for which they were preparing. These letters were submitted to twelve members of the faculty at the Union Theological Seminary who ranked them according to the degree to which the individual's letter indicated "general fitness for the position." To obtain a criterion by which to evaluate these estimates, five of the students' teachers ranked them according to general ability, intelligence, and tact. They were similarly rated by one another — each student ranking the other twenty-four members of the group in these traits. It is interesting to note that the teachers' and the student associates' ratings correspond closely for general ability with a correlation of .90 and for intelligence with a correlation of .83. The correlation for tact is .59. However, the real problem is the extent to which the estimates made from the letters of application agree with the estimates made by teachers or associates who are actually acquainted with the individuals. These correlations are given in Table LXVII. For instance, the correlation between general fitness for this type of work in the opinion of those evaluating the letters and general ability as estimated by the teachers who were in contact with the applicants in question is .56,

TABLE LXVII. CORRELATION BETWEEN ESTIMATES OF GENERAL FITNESS FOR A POSITION BASED ON LETTERS OF APPLICATION AND ESTIMATES OF TRAITS BY ACQUAINTANCES<sup>1</sup>

	TEACHERS	ASSOCIATES	TEACHERS AND ASSOCIATES
General ability.....	.56	.46	.50
Intelligence.....	.58	.44	.44
Tact.....	.20	.18	.22

<sup>1</sup> After Poffenberger and Vartanian.

whereas the correlation between this same estimate from the letters and the judgment of student associates as to general ability is .46. When the judgments of teachers and associates are combined for each individual into a single judgment of general ability, these figures correlate with the estimates from the letters to the extent of .50. Similar correlations are given for the other traits.

It is to be noted that there is a fair correlation between estimates of general fitness based on the letters and estimates by acquaintances as to general ability and as to intelligence. However, the estimates of tact are apparently of no value as far as correlation with the criterion is concerned.

It is interesting to compare the results obtained by pooling the estimates of all the judges with that obtained by the individual judges. If the results for each of the twelve judges are correlated separately with the criterion, combining the three traits into a single figure for a given judge, the average of these correlations is .37. If, on the other hand, the estimates for all the judges are combined in a single figure for each candidate and these pooled estimates are correlated with the criterion, the correlation is .46. This is the same tendency that has been found in other connections, namely, that better results are obtained by combining the estimates of a number of judges than by taking the estimates of any particular judge. The correlation between the criterion and average estimates is usually larger than the average of the correlations between the criterion and individual estimates. This sug-

gests the possibility, if members of a staff are evaluating application letters, of adopting a technique whereby they independently rate the letters and have these ratings then combined into a single figure for each letter.

**Graphology.** A word should perhaps be said regarding graphology in the present connection, because some employers may have the notion that they can infer various character traits from the handwriting of the application letter. Most of the generalizations in this field are based on analogy. It is assumed that writing continuously from letter to letter denotes coherent thought, while writing the letters with breaks between them indicates that the person is addicted to flashes of inspiration; that heavy writing denotes strength of will and persistence; that large bold writing is made by a person with imagination and ambition. These conclusions are not based on empirical evidence. Even such reasoning as that neatness in writing connotes general neatness is unwarranted. Habits are specific rather than general. Ambition to win in golf does not necessarily denote desire to do one's best in the factory. Enthusiasm for social contact at a dance differs from desire to meet people from behind a cashier's window. Neatness in handwriting or in personal appearance is not a universal index of neatness in clerical work.

When the alleged assumptions of graphology are evaluated statistically the results are similar to those found by similar statistical studies of the claims of physiognomy. (Cf. Chapter II.) A group of fraternity members rated one another with respect to the following traits: ambition, pride, bashfulness, force, perseverance, and reserve. Their handwriting was then analyzed with reference to alleged criteria of these traits such as lines sloping uphill or down or fine or heavy lines. (254.) The correlations ranged from .38 to -.20 and two thirds of them were negative. The conclusion was drawn that "there is no indication whatever of these six character traits being betrayed by the signs that graphologists accept. Perhaps there are other signs that do betray character traits, but if so the characterologists are keeping it a secret and careful scientific investigators are unable to discover any." In another instance the judges estimated the intelligence of a large

number of students on the basis of samples of their handwriting in a uniform piece of dictation. The estimates were compared with results in an intelligence test. The correlations for the different judges ranged from .16 to  $-.16$ , indicating utter inability to judge intelligence from handwriting. (428.) It is possible to tell the sex of persons from their writing only about 60 per cent of the time, while, of course, one could guess correctly 50 per cent of the time without looking at the writing at all. It seems that about all that writing manifests is the neatness or perseverance or some other trait involved in actually writing the letter.

**Estimates of one's self.** Inasmuch as the writer of a letter of application often gives some evaluation of his own traits or capacities, it is in point to consider how well one *can* evaluate himself. Studies in which persons have rated themselves in various traits and have had their ratings compared with the ratings of intimate acquaintances reveal the tendencies. (244.) In one instance twenty-five persons ranked one another, themselves included, in a list of traits. For a given trait the average rank assigned an individual by his associates was taken as his actual possession of the trait. It was then possible to note how the rank he assigned himself deviated from this "true" rank. These results are shown in the first column of Table LXVIII, which gives the average of such figures for all the individuals. For purposes of comparison there was also computed the average agreement among the judges in estimating each trait. These figures appear in the second column. In this case a purely chance arrangement would give an average deviation of a little over six steps. The self-estimates in general deviate to almost this extent, while the estimates by acquaintances are appreciably better. The results for another group of individuals who performed a similar experiment are given in the last column of the table. It shows the tendency to overestimate (+) or underestimate (-) one's traits relative to the average estimate of acquaintances. The tendency seems to be to overestimate one's self in the more desirable traits and to underestimate one's self — i.e., give one's self a higher rating than is deserved — in the undesirable traits. Consequently, statements regarding an applicant's mental traits in his own letter, even though sincere, are of dubious value.

## DETERMINANTS OF VOCATIONAL APITUDE 409

TABLE LXVIII. ESTIMATES OF ONE'S SELF COMPARED WITH ESTIMATES BY ONE'S ASSOCIATES<sup>1</sup>

	AVERAGE DIS- PLACEMENT OF SELF-ESTIMATES FROM ESTIMATES OF ASSOCIATES	AVERAGE DEVI- ATIONS OF ESTI- MATES OF ASSO- CIATES	AVERAGE OVER- ESTIMATION OF SELF
Neatness.....	5.8	4.5	+1.8
Intelligence.....	6.0	3.7	+3.0
Humor.....	7.3	4.5	+5.2
Conceit.....	5.7	4.1	-1.7
Beauty.....	6.0	3.8	+0.2
Vulgarity.....	6.1	3.5	-4.2
Snobbishness.....	5.1	4.8	-2.0
Refinement.....	7.2	5.9	+6.3
Sociability.....	5.4	4.7	+2.2
Average.....	6.1	4.4	

<sup>1</sup> From Hollingworth's *Judging Human Character*, by permission of D. Appleton and Company, New York.

### RECOMMENDATIONS

**Difficulties.** Recommendations are required in many employment situations. When a prospective employer does not know the applicant personally, it seems perfectly logical to make inquiry of some one who does. If the former employer to whom inquiry is made is serious and fair and his ability to judge traits accurate, his recommendation should be of some actual value, but there are difficulties with the procedure on these very points. The first of these, about which unfortunately it is impossible to obtain scientific evidence, is the bias or carelessness of the writer of the recommendation. "The recommendation is often sealed with a shrug and opened with a smile." (244.) The writer is often led to overstate the case. He may desire thereby to facilitate the exit of a present employee and to "wish" him on some one else. In such cases the "enthusiasm of the writer may indicate only his joy at a separation long overdue." On the other hand, the writer may be led to understate the case. He may perhaps have had some disagreeable experience with the man so that he wishes him to receive unfavorable

consideration or he may be anxious to keep the employee who is nevertheless seeking other openings and hence misrepresent the man's actual ability in order to prevent some one else from hiring him away. In other instances the writer may have no particular bias, but may simply make inadequate statements of a perfunctory character. Many recommendations are of this sort. The writer feels some doubt as to the value of the whole procedure and as to his ability to evaluate the candidate and simply uses certain set or conventional terms on all such occasions. In such instances the apparently detrimental content of the recommendation does not reflect the applicant's lack of ability, but rather the writer's apathy regarding the applicant's destiny.

Supposing, however, that the writer of the recommendation is unbiased and tries to do his best in evaluating the applicant, there are other possible sources of error and other factors that should be considered before attaching much significance to his statements. Something depends on the aspects in which he is called upon to evaluate the applicant. The preceding chapter showed the necessity for working out carefully the details of a rating procedure and of training the raters if any great value is to be achieved in considering character traits. This has obvious implications regarding the value of estimates made by untrained persons in writing a recommendation. Some aspects, however, may not be as bad as others. It has been pointed out that the more objective traits are rated somewhat more reliably than the subjective. In evaluating a recommendation perhaps more significance should thus be attached to a statement regarding objective than to one regarding subjective traits.

Another thing that should be considered is the relation between the one making the recommendation and the applicant with especial reference to the conditions under which the former has generally observed the latter. If the applicant has been a pupil or parishioner of the recommender, the conditions of observation will have been somewhat different from what they would have been if he were an employee. The conditions under which a trait is judged makes some difference in the reliability of its estimation as the following study shows. (244.) A group of teachers rated one another

## DETERMINANTS OF VOCATIONAL APTITUDE 411

in seven different traits. A group of students rated one another in these same traits and finally a group of students judged their teachers in these traits. In each instance the reliability of each trait was determined by computing the agreement of the judges with each other. The results appear in Table LXIX. The actual

**TABLE LXIX. ORDER OF RELIABILITY OF ESTIMATES OF TRAITS <sup>1</sup>**

	TEACHERS JUDGING TEACHERS	STUDENTS JUDGING STUDENTS	STUDENTS JUDGING TEACHERS
Efficiency.....	1	2	5
Energy.....	2	4	3
Leadership.....	3	3	7
Independence.....	4	1	6
Coöperativeness.....	5	6	4
Cheerfulness.....	7	5	2
Kindliness.....	6	7	1

<sup>1</sup> From Hollingworth's *Judging Human Character*, by permission of D. Appleton and Company, New York.

deviations are not shown, but merely the relative order of reliability for the traits. For instance, with teachers judging teachers efficiency is rated with the most reliability, while with students judging students the most reliable estimates are for independence. There is obviously a fair correspondence between the relative reliability of the traits when teachers judge teachers and when students judge students. The results are quite different when the students are judging the teachers. Some of these reversals are quite understandable. Estimates of kindness and cheerfulness, for instance, are most reliable for the students judging the teachers and much less reliable for teachers judging one another. Kindness is a thing which the students would collectively have a chance to observe in the classroom and the same thing would be true of cheerfulness. Under these circumstances the students would, therefore, make rather uniform judgments of those traits inasmuch as they would observe them operating under the same conditions. The teachers judging one another, however, do not make their judgments under uniform conditions, and one of them will see a man in



a situation where his kindness will manifest itself and another in a situation where there is no such opportunity. On the other hand, leadership is rather poorly judged by the students. It apparently does not manifest itself in the classroom situation. Fellow teachers, however, have rather common criteria in the social environment by which to estimate the leadership of the teacher in question and do so more effectively than students. The point then is that estimates of traits depend for their value to quite an extent upon the relation between the judge and the judged. It would seem offhand that the recommendation of a former employer who had observed the individual in the actual industrial situation would be more valuable than that of a person whose observation had been confined to other situations.

**Kinds of recommendations.** There are three general kinds of recommendations. The first is the testimonial which the applicant solicits and takes with him when leaving an employer. This type of testimonial is usually a brief statement of satisfactory service. It cannot go into much detail nor give anything of a confidential character because the applicant sees the letter himself. A second type of recommendation consists of a letter written directly to the prospective employer at the request of the applicant. This is somewhat better than the former type because it involves confidential material; the previous employer can write with little restraint, and if he cares to do so can give an unbiased account of the individual's qualifications as far as he can judge them. The third type of recommendation consists of a letter in response to an inquiry from a prospective employer to some previous employer. This has the great advantage of calling for, and probably obtaining, the specific information that is wanted. Whereas in the other cases the prospective employer may receive a lot of high-sounding irrelevant material, in this case he obtains information primarily on the points which he considers significant in his particular situation.

The last of these types of recommendation is the only one that is worth serious scientific consideration. The conventional method is to write a simple personal letter of inquiry, but there is the possibility of some refinements in this method. It is feasible, for in-

## DETERMINANTS OF VOCATIONAL APTITUDE 413

stance, to construct the inquiry in such a way as to save a lot of time on the part of the one answering it as well as on that of the one who is subsequently to evaluate it. While it is possible to ask specific questions requiring a more or less detailed answer, a cue may be taken from the general technique of mental tests. Instead of using questions requiring sentences for an answer, one may obtain the same information by having the writer merely indicate his answers by a few check marks or at the most by a few words. In mental tests, for instance, instead of asking the subject to write the opposite of the words, we present several alternatives and he checks the correct one. Exactly this same procedure may frequently be adapted to the recommendation blank. The following blank is typical.

*Dear Sir:*

Mr.....has applied to us for a position as..... and has named you as a former employer. It will help us if, in entire confidence, you will give us the information requested below. We shall be glad to reciprocate at any time.

1. In your opinion is he honest and responsible? Yes.....No.....
2. Is he temperate with tobacco and alcohol? Yes.....No.....
3. Does he possess skill in the work named above?  
High skill.....Generally qualified.....Doubtful.....No skill.....
4. He states that he was in your employ as.....from..... to..... Does this correspond to your record? Yes..... No.....
5. He states that he left because.....  
Is this an adequate statement? Yes.....No.....
6. He states that he received in salary or commission..... per..... Is this correct? Yes.....No.....
7. Would you reemploy him? Yes.....No.....
8. If not will you please give reasons.....
9. If you have further information that will assist us in helping him make the most of his opportunity, kindly indicate it.....
10. If you have further information that can better be given in personal communication with our representative, check here.....

For obtaining information dealing more specifically with various traits a scheme somewhat similar to a rating scale has sometimes

been used. After an introductory statement like the preceding, the reader is requested to check or ring the word in each line that most nearly describes the applicant.

<i>Physical appearance</i>	Commanding	Pleasing	Average	Unattractive	Insignificant
<i>Clothes</i>	Stylish	Well dressed	Ordinary	Untidy	Shabby
<i>Manners</i>	Obtrusive	Friendly	Well mannered	Retiring	Bashful
<i>Ambition</i>	Keenly ambitious	Moderately ambitious	Easily satisfied	Indifferent	Lacking
<i>Application</i>	Exceptionally industrious	Industrious	Performs work assigned	Shiftless	Lazy
<i>Persistence</i>	Very persistent	Determined	Ordinary	Easily discouraged	A quitter
<i>Popularity</i>	Very popular	Good mixer	Average	Exclusive	Unpopular
<i>Parents</i>	Wealthy	Well off	Moderate circumstances	Working people	Poor

The above items are, of course, merely suggestive and would necessarily vary with different occupations. However, recasting the recommendation blank into this form enables the prospective employer to obtain the desired information with a minimum outlay of time on the part of the one filling out the blank and of the one evaluating it. If an individual is repeatedly solicited for recommendations which he must answer at length, he naturally drops into perfunctory habits. If, however, the request is presented in such a fashion that he can in a very few minutes check the answers, he will react to it much more favorably and be more apt to exercise his best judgment.

#### THE INTERVIEW

Employees are seldom hired and probably should not be hired without a personal interview by some member of the staff. In the first place, there is a rather general feeling that it is desirable to see the applicant and talk to him with a view to sizing-up certain traits that might not be revealed by other procedure. In the second place, the interview may give to the applicant information about the nature of the proposed work and about the company so that his subsequent experience will not run counter to his initial impressions. In the third place, the interview affords an opportunity to

make a friend for the company so that the applicant will desire to work for it.

The first of the foregoing functions is the one that has received the greatest stress and experimental study. If the information as to the applicant's mental or other qualifications revealed by the interview is valid, this constitutes, of course, a convenient and expeditious method of hiring. Many executives have, however, a probably unfounded confidence in their ability to predict occupational success by this method. At any rate, the interview is such a common practice that it is desirable to investigate its worth scientifically, particularly from the standpoint of the value of judgments regarding the applicant's qualifications. Interviews vary widely in character. In some cases a rather perfunctory set of questions is asked with a view merely to keeping the applicant engaged while he can be watched. In other cases a more flexible and exhaustive questioning is used with a view to obtaining as much information as possible about the applicant's qualifications.

**Factors making for unreliability.** The customary interview procedure is not all that can be desired, for there are a number of psychological factors that tend to produce unreliability. In the first place, the interviewer is prone to use personal generalizations about such things as physiognomy. (Cf. Chapter II.) If one has, for instance, had an unpleasant experience with some person with a long nose or red hair, he is likely to impute the same unpleasant characteristics to an applicant possessing these same physiognomic aspects. Many persons have some almost unconscious generalizations of this sort which doubtless considerably influence their judgment of people whom they observe. To be sure, such generalizations may occasionally be sound and based on psychological principles, but the difficulty is that it is impossible to ascertain whether or not the generalization is sound unless recourse is had to statistical methods. The interviewer should then note whether he is using generalizations of this sort and should guard against them unless he actually knows that they have a statistical foundation.

A second factor making for unreliability of the interview is the frequent assumption that habits are general rather than specific. It is assumed that a habit formed in one field with reference to one

kind of situation will operate equally in other fields. This tendency was mentioned in connection with the letter of application. A very common assumption is that the applicant who is neat in dress will be likewise neat in work, or that a person who talks rapidly and seems very much alive will be a rapid worker, or that one with awkward physical posture will be inaccurate and clumsy in his manual work. As a matter of fact habits are not usually generalized to this extent, but are more frequently specific in character. A habit involves essentially the forming of a path of low resistance between two portions of the nervous system so that when an impulse reaches one of these portions, it will discharge readily into the other. This tends to make succeeding impulses discharge in the original direction rather than spread out in a general fashion. The habit, for instance, of looking in a mirror and adjusting the necktie deals specifically with the motions involved in adjusting the tie and not with the motions involved in making a micrometer adjustment on machine tools. The neural pathways in the two instances are quite different. Again, the nervous connections that lead to the speech muscles of a rapid talker do not lead to his hands and give a presumption that he will be rapid in manual work. Similarly, a person who is clumsy in the control of his larger muscles, feet, and arms may not be equally clumsy with the fine coordinations of his fingers in doing delicate machine work. It is easy to conceive special cases in which, even if the habit were somewhat generalized, other factors would enter to break it down. An applicant for an executive position might present himself with grimy hands. This might not reflect personal untidiness, but rather the fact that while waiting for a job in his own line he took a mechanical job as the next best thing. The particular traits leading him to make such a shift might really be of the sort that would make him better qualified for the executive position. In other instances a person might be somewhat unkempt because of such tremendous interest in an invention or piece of research that he was conducting that he temporarily neglected his appearance.

A third factor that contributes to the unreliability of the interview is the "nervousness" of the applicant. It is quite possible that many an applicant, in the excitement of the situation, par-

ticularly if it is a very important matter for him, will be in a distinctly abnormal mental condition. An individual who is usually fairly calm may under these circumstances show what seems like distinct nervous instability. In giving tests it will be recalled that a "shock-absorber" test often precedes the tests proper to alleviate this initial emotional disturbance. A skillful interviewer will probably be able to determine in the course of the conversation whether or not the applicant is in such a state and in the former instance should be able to remedy the condition.

**Demonstration of unreliability.** While the foregoing factors presumably make for unreliability of the interview, it is further possible to study the matter statistically in the same fashion that reliability has been studied in other connections. Fifty-seven applicants for sales positions were interviewed individually by twelve different sales managers. (244, 65.) These managers were allowed to conduct the interviews in whatever fashion they wished, but at the conclusion were required to rate the individual as to "suitability for the position in question." These ratings were then cast into such form that it was possible to assign each applicant a rank from 1 to 57 for each judge. The results for a typical group of applicants are given in Table LXX. It will be seen that there is a marked disagreement among the interviewers. Applicant C, for instance, is placed first by one interviewer and fifty-third by another. Several other applicants show discrepancies in the ratings of about this magnitude. It is to be remembered that these interviewers were sales managers of considerable experience in making such judgments and the extent of their disagreement in rating the same applicants is rather disquieting.

In another instance six sales managers interviewed thirty-six applicants for sales positions. (521.) The results may be summarized in a word. With twenty-eight of the thirty-six applicants the managers disagreed as to whether the individual should be in the upper or the lower half of the group.

Another similar study was made in employing truck salesmen. (541.) A want ad was inserted in the paper and on the basis of the letters received twelve applicants selected. They were interviewed individually by six sales managers and also by a psycholo-

TABLE LXX. RANKS ASSIGNED APPLICANTS BY SALES MANAGERS WHO INTERVIEWED THEM<sup>1</sup>

APPLICANT	SALES MANAGERS											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
A	33	46	6	56	26	32	12	38	23	22	22	9
B	36	50	43	17	51	47	38	20	38	55	39	9
C	53	10	6	21	16	9	20	2	57	28	1	26
D	44	25	13	48	7	8	43	11	17	12	20	9
E	54	41	33	19	28	48	8	10	56	8	19	26
F	18	13	13	8	11	15	15	21	32	18	25	9
G	33	2	13	16	28	46	46	32	55	4	16	9
H	13	40	6	24	51	49	49	52	54	29	21	53
I	2	36	6	23	11	7	7	17	6	5	6	9
J	43	11	13	11	37	40	40	46	25	15	29	1

<sup>1</sup> From Hollingworth's *Judging Human Character*, by permission of D. Appleton and Company, New York.

gist as to fitness for the position. There was a fair agreement among the interviewers as to the two best and the two worst candidates. In the other cases, however, the agreement was very little, indeed. The average deviation of the judges was a trifle over three places, and inasmuch as there were only twelve possible places this deviation was very serious. After the estimates were pooled to get a consensus of opinion, it was possible to estimate the reliability of any individual judge by correlating his rating with the consensus. The correlations for the seven judges were as follows: .12, .38, .01, .72, .58, .47, and .71. The last of these is the correlation for the psychologist. He did practically as well as any of the experienced sales managers.

**Improvements in technique.** The foregoing indicates some of the difficulties and the unreliability involved when an interviewer is left to his own devices. This points to the desirability of making, if possible, some improvement in the technique of the interview or else of discarding it altogether. We have seen in other connections that if judgments are to be made a better result can usually be obtained by pooling the estimates of several judges than can be ob-

tained by an individual judge. The obvious implication is the desirability of employing more interviewers. If the applicant is interviewed by several members of the staff and their judgments are pooled, there should theoretically be an increase in the value of the procedure.

Another desirable feature of the interview is the establishment of *rapport* at the outset. Just as in mental test procedure it is desirable to get the individual into the proper attitude so that he will coöperate and do his best, it is likewise necessary in the interview to get him into this proper attitude. This requires tact and skill on the part of the interviewer. It is desirable to get the applicant's confidence and make him feel that effort is being made to help him as well as help the company. Given this initial *rapport*, the procedure will probably go more smoothly and reliably as a result.

A third feature is the use of crucial questions the value of which is actually known. An analysis of the job and the job specifications will often indicate the type of information that is most valuable in a given case. Knowing what is needed, the interviewer can then go after that definite information which will be relevant. Care should be taken, however, that the questioning is not too cursory, but sufficiently flexible so that the applicant will reveal his own characteristics.

A final suggestion for improving the interview technique consists of having in mind during the interview certain specific traits which are to be observed. If, for instance, the interviewer is watching for such things as appearance, manner, energy, coöperativeness, and confidence, he may continually have these things before him during the conversation. Some concerns go so far as to have a blank for recording "interviewer's impressions." Traits such as the foregoing are listed on this blank and the interviewer makes notes covering these definite points during the conversation. This procedure has the advantage of directing the attention to certain specific things and getting away from mere general impressions to a critical evaluation of definite traits.

Rating scale for interviewers. Instead of merely having before him a list of traits as a guide while making his estimation, it is possible for the interviewer actually to use in systematic fashion a



rating scale similar to that described in the preceding chapter. The details of the development of a rating scale will not be repeated, but it is possible to adapt the technique of the ordinary rating scale to the conditions of the interview. In order to make actual ratings, to be sure, it is often necessary to know the individual and to have observed him for some time. Some traits, however, may manifest themselves to a certain extent at first sight. Moreover, employment men often find it necessary to make at least some sort of an estimate of a man in a preliminary interview. This estimate may be better than nothing and whatever may be done to increase its reliability is desirable. The actual traits to be estimated in any given case will depend on the local situation and the nature of the vocation and whether they are of a sort that can be judged without long acquaintance.

The following procedure is typical of the man-to-man scale used for an employment interview. The interviewer is first of all provided with a rating scale blank on which he is to make up his master scale after the fashion described in Chapter XII.

#### INTERVIEWER'S RATING SCALE FOR EXECUTIVES

Make up a list of twenty-five or more executives whom you know very well. Include in this list some who rank very high, some who are intermediate, and some very low in traits such as appearance, energy, social attitude, tact, and initiative. Be sure that your preliminary list is representative.

*Appearance and manner.* Disregard every characteristic of the executive except the way he will impress people by his physical bearing, neatness, and facial expression. Consider whether he will be repulsive or whether he will fall somewhere between the extremes. Select from your list the man who ranks the highest in this respect and note his name on the first line which is marked "Highest Mr. ...." Then select the one who ranks lowest in appearance and manner and put him on the bottom line. Then, still considering only this same factor, select a man on your list who falls midway between highest and lowest, indicating him on the middle line. Then determine one who ranks between the highest and the middle and another who ranks between the middle and the lowest indicating their names on the appropriate lines.

Highest...Mr.....	20
High....Mr.....	16
Middle....Mr.....	12

## DETERMINANTS OF VOCATIONAL APTITUDE 421

Low.....Mr.....	8
Lowest....Mr.....	4

*Energy.* Consider the executives on your list from the standpoint of energy, that is, the way in which they actually go at their work. They will probably range all the way from listless up through the type that gets things done to the one who is full of energy or is a "live wire." Considering only this characteristic of energy select the one man from your list who possesses it to the greatest degree and put his name on the first line. Similarly select the one who ranks the worst in this respect and list him on the last line. Fill in the names for the others just as you did with reference to the preceding trait.

Highest...Mr.....	20
High.....Mr.....	16
Middle....Mr.....	12
Low.....Mr.....	8
Lowest....Mr.....	4

*Social attitude.* Consider each of your men from the standpoint of how he will act when meeting people in a business way, whether he will be formal and constrained or meet people halfway and manifest a breezy informality. On the basis of this trait alone select five of your men in the preceding fashion and list them on the appropriate lines.

Highest...Mr.....	20
High.....Mr.....	16
Middle....Mr.....	12
Low.....Mr.....	8
Lowest....Mr.....	4

*Tact.* Consider them from the standpoint of their ability to get along harmoniously with other individuals. Observe whether they antagonize, make only occasional breaks or are extremely tactful and harmonious. On the basis of tact select five of your men as in the preceding cases and enter their names on the following lines:

Highest...Mr.....	20
High.....Mr.....	16
Middle....Mr.....	12
Low.....Mr.....	8
Lowest....Mr.....	4

*Initiative.* Consider their tendency to get things done in the face of obstacles or opposition. They may be very meek and irresolute or they may possess various degrees of stick-to-itiveness. Bearing in mind this trait alone select 5 men in the preceding fashion and list them on the following lines:

Highest... Mr.....	20
High..... Mr.....	16
Middle... Mr.....	12
Low..... Mr.....	8
Lowest.... Mr.....	4

The interviewer at his leisure fills out a blank similar to the above. He now has his master scale by which he may evaluate the individual in the interview. The technique then consists of having this master scale before him while conducting the interview and actually comparing the applicant with the various men on the scale in the different traits there listed. If the applicant with reference to appearance, for instance, impresses him as similar to the first man listed on the master scale, he assigns him a value of 20. A separate "interviewer's rating blank" is, of course, provided for recording these judgments.

The interviewer may make these judgments during the conversation and record them as he forms them or he may possibly hold them in memory till the conclusion of the interview and note them shortly thereafter. In general it would seem better to make some notations even during the interview. Certain acts of the applicant may indicate a marked presence or absence of some trait which would perhaps be forgotten before the end of the conference. The essential point, however, is that the interviewer has before him this concrete master scale with which he is comparing the applicant man-to-man while he is talking to him.

The method of rating by defined groups may likewise be used in this connection. A typical blank might read as follows:

#### INTERVIEWER'S RATING SCALE FOR EXECUTIVES

Imagine all the executives of this type whom you have ever known divided into five classes of equal size on the basis of each of the traits listed below. Have this blank before you and keep the classes in mind during the interview. Try to compare the applicant with these other groups and determine in which he should be located. Consider one trait at a time. If you cannot reach a decision regarding a certain one, pass on to the others and return to it later in the interview. When you have come to a decision regarding a certain trait, check in the appropriate column. You may grade as finely as you wish by placing the check toward the right or left of a col-

## DETERMINANTS OF VOCATIONAL APTITUDE 423

umn according as you consider that the applicant stands high or low in a given group.

	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
<i>Appearance and manner.</i> How he will impress people by his physique, bearing, neatness and facial expression	.....	.....	.....	.....	.....
<i>Energy.</i> Whether lary and listless, gets things done or is actual live wire.	.....	.....	.....	.....	.....
<i>Social attitude.</i> Whether meets people formally or halfway and informally	.....	.....	.....	.....	.....
<i>Tact.</i> How he gets along with people	.....	.....	.....	.....	.....
<i>Initiative.</i> Tendency to stick and get things done in the face of opposition	.....	.....	.....	.....	.....

In similar fashion the graphic rating scale may be adapted to use during the interview. A typical blank might be as follows:

### INTERVIEWER'S RATING SCALE FOR EXECUTIVES

During the interview have in mind the traits listed below. Try to observe the applicant as to the extent to which he possesses these various traits and indicate by a check mark on each line your judgment of the applicant. Be careful to judge each trait independently of the others. If it is feasible to make these judgments during the interview, do so, although it may be desirable to postpone some of them until the end. After the interview, go over the results again immediately, for you may wish to make some slight revision.

<i>Appearance and manner.</i> Consider how he will impress persons by his physique, bearing, neatness and facial expression.	Repulsive	Unimpressive	Satisfactory	Noticeable	Excites admiration
--	-----------	--------------	--------------	------------	--------------------

**Energy.** Consider the way he will presumably go at his work.

Full of "pep" "live wire"	Active	Will get things done	Half-hearted	Lazy and listless
------------------------------	--------	----------------------	--------------	-------------------

**Social attitude.** Consider how he will act when meeting people in business way.

Formal and constrained	Somewhat reserved	Will meet halfway	Cordial	Breezy and informal
------------------------	-------------------	-------------------	---------	---------------------

**Tact.** Consider his ability to get along harmoniously with others.

Very tactful	Will seldom make a break	Will make occasional mistakes	Indiscreet	Antagonizing
--------------	--------------------------	-------------------------------	------------	--------------

**Initiative.** Consider his tendency to get things done in the face of obstacles.

Meek	Irresolute	Moderate stick-to-it-iveness	Surmounts most obstacles	Very persistent
------	------------	------------------------------	--------------------------	-----------------

With blanks of this sort the interviewer can more adequately record his judgment of the applicant. The procedure of scoring and weighting the items is exactly the same as that given in the discussion of rating scales.<sup>1</sup>

This rating scale procedure probably represents the best contribution that psychologists have made to date to the technique of conducting the interview. In the light of the general unreliability of estimates of traits and the danger that various factors will influence judgment, any efforts to put the interview on a more scientific basis are worth considering. The rating scale technique which has proved of some value in passing judgment upon present employees may likewise contribute something to the improvement of methods of hiring persons on the basis of a personal interview.

While the foregoing discussion of the interview has centered upon the question of obtaining information about the applicant, this should not be construed as minimizing its other functions. It should give, as well as obtain, information. Employment is not merely a process of selection — it should be a mutual process. The company is entitled, of course, to information about the applicant's qualifications, but the applicant likewise is entitled to information about the company and the proposed job. Many a man takes a job under false pretenses on the part of the company. He

<sup>1</sup> For a graphic rating scale for interviewing salesman, cf. 278, 203.

assumes that it is a stepping-stone to other work, but finds later that it is a blind alley. He is shocked to find that it has much more dirt or much greater hazards, or is more irksome than he had anticipated. He supposed that his duties would be mainly inside the building, but discovers later that he has to go out on the road in cold weather. Probably it does not occur to him at the time of the interview to inquire regarding these matters. However, the subsequent development at variance with his assumptions produces a dissatisfied employee or even a "quit." This could have been obviated by foresight on the part of the interviewer. The applicant is questioned, tested, rated, recommended, and analyzed, and he is entitled to some reciprocal information. In the interest of ulterior satisfaction and harmony the interviewer should put the cards on the table and tell the applicant about all aspects of the proposed job that are of possible significance. Even though the applicant at the moment merely wants a job and a pay envelope, he should be compelled to enter the job with his eyes open, knowing its disadvantages as well as its advantages.

In addition to giving information to the applicant the interviewer should strive to make a friend for the company. The traditional importance of first impression is applicable in this connection and the first impression of the company which the man gets at short range is usually in the employment office. If the interviewer takes the proper attitude, and tries to interpret to the applicant the policies and ideals of the company, this may be the beginning of a permanent friendly relation. Effort may be definitely made to "sell" the company to the applicant. If he is thus sold and his attitude is firmly ingrained, this will often bridge some inevitable vicissitude in the industrial cycle and the employee will stick with the company and support it until matters are adjusted. Even if an applicant is rejected, but goes away with the feeling that this would be a fine place to work and transmits this feeling to his acquaintances, he constitutes an asset rather than a liability.

Thus the employment interview has these three distinct functions. It obtains information about the applicant which may supplement other data, it gives to the applicant information about the company and the policies, and it makes a friend for the com-

pany. The ideal interviewer should not be merely an examiner or inquisitor, but also an instructor and a salesman.

#### SUMMARY

While the technique of mental tests for employment purposes is preferable to the use of less objective indications of vocational aptitude, there are situations in which these latter deserve consideration. They may sometimes prove a valuable supplement to tests. The more variables investigated, the greater the probability of finding some with high correlations with the criterion. Moreover, the best variables from the predictive standpoint are those which have low intercorrelations — i.e., involve discrete rather than overlapping factors; and it sometimes proves that the tests intercorrelate rather highly while the miscellaneous factors have low correlations with the tests. In this case the latter may well be embodied in the regression equation. In instances where test technique is not feasible at all, it is worth while to investigate these miscellaneous factors and determine their value so that they can be used more systematically than heretofore.

Academic record in school or college has some predictive value. The school curriculum itself is a selective process inasmuch as the intelligent pupils progress more rapidly than the unintelligent. Statistical studies show that early marks in school are rather prognostic of later school marks. Academic achievement, moreover, seems related to the type of success that causes a listing in *Who's Who*, marks in technical school have been shown to bear some relation to subsequent salary, and grades at West Point have been shown to be somewhat prognostic of military success. The amount of education obtained by the applicant may show whether he has mastered certain fundamentals which he will need in his vocation, and his rate of progress as shown by age and grade at leaving school is an indirect indication of his general intellectual capacity. Efficiency in some types of work, such as clerical, has shown appreciable correlations with years of schooling. Each situation, however, must be investigated for itself because equivocal results have been found. Achievement in special school subjects, such as manual training, has rather obvious implications regarding aptitude for

similar work. The choice of certain subjects in an elective curriculum gives some index of a person's interests and perhaps also of his abilities. However, school marks at best are distinctly inferior to tests where the latter are feasible. A three-hour test was far more predictive of the first two years' achievement in an engineering college than was the entire high school record.

It is sometimes possible from a person's early proficiency in a given occupation to predict his subsequent success therein. With salesmen first-year production seemed a rather good index of production in following years. In a machine shop there proved to be a relation between the accidents encountered by an employee in successive quarters of the year. In some clerical jobs there was considerable relation between efficiency in successive months especially with reference to speed.

The items on personal history or application blanks have been analyzed as to their vocational significance. The technique consists of tabulating various biographical items for different occupational groups and noting which items are differential. The implications of height and weight for heavy muscular work are obvious. A more subtle problem is the relation of stature to selling ability. Salesmen as a whole appear appreciably larger than the average man, but within a given selling organization it is sometimes found that the largest men are the best producers and sometimes, perhaps more frequently, that the best salesmen are those who are well above the average stature, but not of extremely large proportions. The age of the employee has been found of significance with some kinds of clerical workers and especially with salesmen. It appears that men of middle age are more effective in this latter capacity even when allowance is made for the effect of experience. There has also been found a relation between age and turnover. In the early years we have the natural instability of youth seeking a vocational objective. Later, perhaps in the thirties, there is greater stability while raising families and buying homes. Then when domestic responsibilities lighten there comes a frequent search for one's ultimate vocation. This produces some instability up till perhaps fifty, when interests have become fixed and a profitable change in employment is unlikely. The common notion that



married employees are superior seems to have some statistical foundation. Likewise other dependents appear to afford an additional motive. It seemed that a salesman with one or two children was more effective than a man with more or less than this number. Previous experience may be of vocational significance either from the standpoint of its amount or its nature. It is not universally true, however, that the more experience the better, for other factors may bring a man with long experience to the employment office. The type of previous vocation may be of interest in ascertaining the most profitable sources of supply for a given occupation. It is possible with these various items of personal history to state certain minimum qualifications in each respect. It is better to weight the items so that they can be combined into a differential score. This may be done arbitrarily or sometimes according to a regression equation. In such cases a better prediction can be made than by attempting to evaluate the items separately.

The letter of application has been studied with reference to its reliability and validity in indicating certain traits or general fitness for a position. Different judges disagree tremendously in estimating traits from a letter, while the same judge agrees with himself in subsequent estimates to only a fair degree. Estimates as to general fitness for a position give only fair correlations with estimates of some traits made by acquaintances and with other traits the correlations are negligible. However, the situation is improved by pooling the results of the judges regarding a given letter before comparing them with a criterion. This suggests that, if application letters are to be used at all, the most satisfactory procedure would be to have several members of the staff evaluate them independently and combine their judgments. Generalizations as to character traits manifested by handwriting are without statistical foundation and have a certain amount of actual statistical refutation. Little significance should be attached to an applicant's own evaluation of his personality traits. Such evaluation has been shown to be very unreliable and to involve usually too good an opinion of one's self.

Recommendations are often worthless because of the prejudice or carelessness of the writer. When this is not the case, note should be taken of the greater reliability of estimates of objective traits as

compared with subjective. Moreover, the estimate depends on the conditions under which the recommender has observed the applicant. The testimonial and the letter written at the request of the applicant to the prospective employer are of little value. The best form of recommendation is an answer to a specific inquiry from the prospective employer because attention is centered on the particular information that will be of value. The technique of making inquiry may be improved somewhat by arranging a blank so that the recommender has merely to choose certain alternative answers to questions or check in certain spaces. This saves the time of the one filling out the blank and he is more apt to do it seriously and in a less perfunctory manner. It also gets very specific and unequivocal information.

The employment interview has certain shortcomings. The interviewer is prone to use personal physiognomic generalizations, to assume that a certain habit as manifested in one's appearance is general and will apply to his work on the job, and to misinterpret excitement in the interview situation as characteristic of the applicant elsewhere. When different members of the staff interview a group of applicants independently and then compare their estimates, the findings are rather disquieting. This was particularly the case with sales managers interviewing applicants for selling positions. Suggested improvements in the interview technique are the use of more interviewers, the establishment of *rapport* at the outset, and the use of crucial questions the significance of which has been established. Finally, if the applicant is to be rated with reference to specific traits, it is possible to adapt the rating scale technique so that it can be used during the actual interview. The man-to-man scale, the method of defined groups, and the graphic rating scale can all be used for certain traits that manifest themselves without long acquaintance. With this method it is possible to rate the applicant during the interview on crucial points, to get away from the halo of general impression, and in general to obtain results that have greater reliability and validity.

In addition to obtaining information about the applicant the interview serves two other important functions. It gives the applicant information about the proposed work so that he enters with

his eyes open as to its nature, as to the working conditions, and the possibilities for advancement. There will then be no discrepancies between his expectations and the ultimate facts so he will not be dissatisfied. The interview also affords an opportunity to sell the company to the applicant and to make a friend. The interviewer should not merely be an examiner, but an instructor and a salesman.

## CHAPTER XIV

### TRADE TESTS

#### TRADE TESTS VS. TESTS OF INNATE CAPACITY

THE distinction has already been made between tests of capacity and tests of proficiency. In the case of the former we are concerned with innate aspects of the individual — certain potentialities which he possesses and which may be indicative of his subsequent success in the job. We use such tests, for instance, to determine whether he has the proper attention and reaction time to make a good tire-builder — a job which he has never tried before. In the case of tests of proficiency we are interested merely in certain acquired abilities which he possesses at the present time. We want to know what sort of carpenter or plumber he is when he enters the employment office. In the former instance we are concerned with predicting what he will ultimately be able to do and with measuring a sample of his innate capacity which will enable us to make such a prediction. In the latter we are interested in some particular ability or skill which he possesses now. In the latter instance, obviously, we are not attempting to prophesy, but merely to determine present conditions.

#### NEED FOR TRADE TESTS

The need for such tests of proficiency arises in industry when hiring a person who is presumed to have a certain amount of trade experience. Machinists, carpenters, electricians, and the like apply for a job on the basis of their previous experience in that field. They frequently carry a journeyman's card or state that they have served for a certain length of time as an apprentice in an establishment. Trade tests are designed to supplement this information. It is often undesirable to take the applicant's own statement as to his proficiency or to take his card at its face value. This fact was brought home to psychologists in rather vigorous fashion by the

experience of the Committee on Classification of Personnel during the World War. Something like half the personnel of the army was engaged in duties of a specialized trade character. It was obviously desirable to assign to those duties soldiers who had functioned in a similar capacity in civil life. If a given unit contained a man who had previously been a barber and another who had previously been a plumber, there was obvious economy in giving them that same type of work to do in the army rather than having the plumber cut hair and the barber mend the leak. Efforts were made to determine occupational status in interviewing recruits, but it developed that such interviews were unsatisfactory. On the average, of the men who professed trade ability in the interview 6 per cent actually proved to be experts, 24 per cent journeymen, 40 per cent apprentices, while 30 per cent were novices. In other words, approximately one third of the recruits who claimed that they were carpenters could not drive a nail and one third of the self-styled automobile mechanics did not know a spark plug from a carburetor. Hence it became imperative for the committee working on these problems to develop some means for determining objectively a man's trade ability regardless of his own statement of his qualifications. In the army were developed the first extensive trade tests and the technique used on that occasion is of rather general applicability.

This problem of trade tests, to be sure, is not of the magnitude of that previously discussed in connection with tests of innate capacity. The present trend in industry is toward a subdivision of labor so that a given worker performs only a relatively minor operation. Whereas formerly one person made the entire shoe, now one man cuts the sole, another cuts the upper, another stitches them together, and another puts on the heel, so that the trade of shoemaker is becoming extinct. In one large concern, for instance, there are 4000 people working on tools which are necessary for the automatic machines run by 15,000 others. Each man works, moreover, on only one machine so that there is no need as formerly for all-round toolmakers. There are still, however, many situations that require persons who actually have some trade proficiency. This is especially true of smaller concerns where the operations have

not been so minutely divided, but even in large organizations there is need for plumbers, carpenters, lathe operators, truck drivers, electric wirers, and the like. Hence the test method of determining the actual trade ability of a prospective employee has considerable applicability.

#### REQUIREMENTS OF TRADE TESTS

**Administration by examiner with no trade knowledge.** There are several requirements which a trade test must meet if it is to fulfill adequately its purpose in the practical situation. In the first place, it must be so constructed that it can be administered by an examiner who has little or no knowledge of the trade in question. This need arises because of the frequent desirability of a centralized process of hiring. A large employment office is often so organized as to do all the employment without consulting specific foremen or other members of the factory staff regarding individual applicants. Hence it would be almost impossible for the examiners to be completely familiar with all the trades involved. Even if the procedure were decentralized, there would be no guarantee that the foreman would administer all the tests in the same fashion. It is obviously fundamental from the scientific standpoint to give every applicant exactly the same test procedure.

**Score independent of examiner's judgment.** In the second place, the tests should be so constructed as to yield a rating independent of the examiner's judgment. The score should be entirely objective and quantitative. This point is related to the preceding. If it were necessary to take an iron hook made by an alleged journeyman blacksmith and rate it as excellent, good, average, fair, or poor, it is probable that there would be marked disagreement between examiners. One of them might notice whether the ring at one end was perfectly round, another might be more concerned with the shape of the point, while a third might note especially whether the general dimensions conformed to specifications. If the hook happened to be well made in one of these respects but not in the others, the applicant's rating would depend largely on who rated his test. It is possible, however, to devise tests in such a form that they can be given by a person with no trade knowledge and never-

theless yield the same result for a given applicant regardless of who has been responsible for his examination and score.

#### PRINCIPLES ON WHICH TRADE TESTS ARE BASED

There are two principles according to which test material may be constructed. A person who is successful in a trade possesses, on the one hand, a certain amount of skill and, on the other, a certain amount of information. A machinist, for instance, is able to set the chisel and operate the feeds on a lathe. He also has certain information about a lathe and can tell the difference between the head-stock and the tail-stock. In attempting to determine whether he has had experience in lathe work we have then two possible avenues of approach. We may ascertain through some standard performance just how well he can manipulate the parts or we may find out how much information about the machinery and materials he has acquired. In general a skilled tradesman will be able to give a good account of himself either in actual performance or in answering questions dealing with his work.

#### KINDS OF TRADE TESTS

**Oral.** The different kinds of trade tests that have been used fall into four general classes: oral tests, picture tests, written tests, and performance tests. Illustrations of each type will be given later. In the oral test the examiner asks the questions verbally and the applicant replies in the same fashion. The questions deal with tools, materials, processes, or other information that a tradesman would be apt to have at his command. The oral test was used quite extensively in the army. The examiner and the soldier sat in a small booth at opposite sides of a table. The examiner was provided with blanks from which he asked the questions verbatim and on which he wrote the soldier's replies. The blanks also contained the correct answers to each question with the amount of credit to be given for each.

**Picture.** In the picture trade test the applicant is questioned regarding the details in pictures of machinery or tools used in the trade. For administering the test two folders are usually provided, one for the examiner and one for the applicant. The latter con-

tains the pictures numbered appropriately. The examiner's folder contains the questions similarly numbered as well as the answers with appropriate credit for each.

**Written.** The written trade test is somewhat similar to the oral except that it is designed for group administration. This necessitates making it sufficiently fool-proof so that the subject can adequately respond by writing or making check marks. In developing the written test a cue was taken from the procedure of the conventional mental tests in having the answers of the multiple choice rather than the single answer form. It will be recalled that in the history of mental tests the earlier types involved the writing of words or phrases as an answer to each item. These answers then had to be evaluated qualitatively. The subsequent trend was toward a series of words one of which was to be selected, thus making the score quantitative and unequivocal. Inasmuch as the multiple choice form of response demonstrated its value in other fields, it has been adopted rather largely in the written trade tests.

**Performance.** In the performance test the subject goes through some standardized typical operation which can be scored according to how he does it or by evaluating the finished product. In the scoring we are not concerned, as in the other types of trade test, with whether the subject gives a certain answer or fails to do so. It is rather a matter of a complex operation or product that must be evaluated. We may, in the first place, consider simply the process which the subject uses in taking the test. In a performance test for a truck driver we would observe primarily the way he handles the truck in going through prescribed maneuvers. In the second place, we may consider the product. A blacksmith may be required to reproduce a piece of stock like a sample and is then graded on the finished product according to how well he actually makes the reproduction. In the third place, we may consider the time consumed in making the product. In general practice these three methods are seldom discrete; two or three of them are combined. We may, for instance, use a process-time test in which the man is required to change the set-up of a lathe in order to do a different job and is scored according to the steps taken in making the change and also the length of time consumed. We may then per-



haps establish a critical score on the basis of both performance and time. Similarly, we may use a product-time test. A typist is given a piece of copy, the finished product is evaluated, and note is made of the length of time taken to complete the work. It is possible also to use a process-product-time test in which all three items are considered. If, for instance, we give a garage mechanic a radiator to repair, we may consider the method he uses, the final job when completed, and the time consumed. Probably the type of performance test most generally used is the product-time test. Its advantage over the process test lies in the fact that it can be scored at leisure and this can likewise be done without any expert knowledge on the part of the scorer. There are situations, however, in which the process-time test is more satisfactory. In most instances, at any rate, the time is taken into consideration.

**Relative merits.** Certain things are to be said for each of these types of test. The oral test has the advantages that characterize any individual as contrasted with any group test. There is the possibility that the subject will misunderstand some trivial point and this can be immediately detected by the examiner. If the subject has any difficulty in making himself clear, he can do so more effectively in conversation. Moreover, the subject may manifest certain reactions extrinsic to the test, such as emotional instability, that will be of vocational significance, and this "clinical aspect" is present in the oral procedure, but missing in the group test.

The picture test is likewise usually conducted orally so that it has the foregoing advantages. It has certain other features that are possibly desirable. It approaches a little more closely to the actual job situation. It is a bit more tangible to look at a picture of a machine tool than merely to talk about it. It gives the tradesman a little more confidence in the test, as it appears more concrete and apparently more practical. It admits, too, of more intricate questions because certain parts of the picture may be lettered and questions asked about those more minute portions which it would be difficult to describe adequately in an oral test. It is also possible that a picture will help the subject to recall further facts because it will be associated with various things in his work and get him

into the proper context. On the other hand, there are disadvantages in the picture test. It is somewhat more difficult to construct and more expensive inasmuch as it involves printing pictures. There is again the danger that the picture will be slightly atypical of the machine with which the worker is familiar. A man who is used to a lathe driven by an independent motor may be a trifle confused when shown a picture of a lathe driven by a belt from a main power line. This slight confusion may be enough to mislead him on lathe questions. Some individuals, particularly those who are of low intelligence, are probably unaccustomed to interpreting pictures anyway, and if they are to be reached at all in this manner they will have to be confronted with the actual machine rather than with a picture of it. Finally, if there are several questions about one picture and the applicant fails to recognize the picture at all, he receives undue penalty because he will fail on all questions dealing with that picture.

While the written test lacks the clinical advantages of the oral, it makes for much more rapid testing just as is the case with any group test. The multiple-choice form of response has likewise the usual advantages. In the first place, the subject does not have to phrase his own answer. Certain individuals with poor ability in grammar might make a rather unfavorable showing, although they were good tradesmen. In the second place, there is no doubt as to the correctness or incorrectness of an answer. The person scoring the blank does not have to judge or subjectively evaluate an item. To be sure, the ordinary questions may be so selected that only a single correct answer seems possible, but even then there is always the possibility of an unsuspected answer that deserves some credit or gives some indication that the subject knows the matter called for. When it is a matter of simply selecting one of several alternatives which are sufficiently discrete, there can be no question as to whether or not the subject deserves credit. In the third place, this type of test may be scored by any ordinary clerk who is unfamiliar with the occupation in question. Finally, the multiple-choice form makes possible more rapid scoring by the use of a stencil which can be aligned over the blank and which will enable incorrect answers to be very readily located.

The performance test has the advantage in some instances that it deals with actual trade skill rather than with information. It is possible for a person to work at a trade and pick up the information without acquiring the requisite skill. This is probably more likely to occur than the opposite tendency to acquire the skill without the information. The skill test is in a way a more direct approach to the ability that is in question. In typewriting, for instance, it is more important to operate the machine effectively than to know the names of the different parts or the adjustments or the difference between various kinds of machines. The performance test is usually more difficult to arrange than the oral or written test. It requires a certain amount of equipment and very often materials that are used up in the process of taking the test. If sufficiently complicated so that fairly elaborate equipment is required, it must be restricted to an individual test. Similarly, if it is scored as a process test, it requires one examiner per subject, and consequently if given on a large scale the organization is more expensive than in cases where the group method is possible.

#### METHODS OF DEVELOPING AND STANDARDIZING

Differ from methods of developing capacity tests. The methods of developing and standardizing or calibrating trade tests are rather similar for all four kinds. For purpose of illustration the method will be presented in detail only for the written test, but the technique to be described is quite typical of methods of evaluating the other tests as well. A somewhat different approach is necessary to these problems from that employed with tests of innate capacity. In that case we were concerned with a group of separate tests such as those for attention, memory, or decision, each composed of many items. The total score for a test was then correlated with the criterion and this was then done for each test separately in order to determine their relative importance and to combine them ultimately into a single score. In a trade test, however, all the items may be of one sort — for example, items dealing with trade information — so that it is not possible to evaluate separately a number of different tests, each composed of many items. Consequently, it is necessary to analyze the individual items and look for

those which are most differential of trade ability. This procedure does not lend itself readily to the computation of correlation coefficients. It is the usual practice simply to take a few outstanding groups of subjects of known trade ability and see which particular items or questions differentiate these groups.

The criterion that is generally used in such procedure involves grouping the subjects according to the ordinary trade classification of novice, apprentice, journeyman, and expert. Most trades have some more or less definite standards of their own regarding these classes. In the army trade tests an expert was defined as "a man with a high degree of trade ability qualifying him for assignment requiring superior workmanship"; a journeyman as "a man with enough trade ability to qualify him for assignment to work which must be done quickly and well"; an apprentice as "a beginner or man with only enough trade ability to make him useful as a member of a group under supervision, not qualified to work without supervision, or where speed and accuracy are prime factors"; and a novice as "a man with no trade ability or so little that he should not be considered when making assignments." When the subjects have been so classified, it is then possible to take each individual question and determine what per cent of the experts answer it correctly, what per cent of the journeymen, what per cent of the apprentices, and what per cent of the novices. If these percentages decrease in the above order there is some indication that this item is differential of the trade ability in question. This method will be described in detail below.

The procedure of selecting and standardizing trade questions in this fashion will be illustrated in the case of a test of the information type for lathe operators. This particular test was devised so that the answers could be written using the multiple-choice type of response. This adapted it to the group method. However, the technique involved in developing it would be equally applicable to the other types of information test and to some extent to the performance test as well.

A preliminary selection of items is the first step in the development of such a trade test. If the psychologist engaged in developing these methods knows little or nothing about the trade in ques-

tion, it is advisable for him either to consult trade journals or to discuss the matter with foremen before devising the test items. In most instances the latter procedure is followed. If the nature of the project is made clear and the principle of finding questions which the good tradesman can answer and the poor tradesman cannot is explained, the average foreman will see what is wanted. He can then be asked to suggest a preliminary set of questions.

These preliminary questions must be worked over carefully before they are in satisfactory final form. Many of those originally devised will be found to be very indefinite or equivocal with the possibility that they will evoke responses of varying degrees of satisfactoriness. It is not feasible to give the foreman a course on the construction of trade test questions. It is better to take his initial attempt and show him where improvement can be made.

The following typical original questions were obtained from a foreman who supervised engine lathe operators. The general program was explained to him and he was requested to submit questions which might be used to determine whether a workman had the requisite information regarding the trade:

1. What is an engine lathe?
2. What is a lathe dog?
3. How fast should a belt run?
4. What is the most vital feature of a lathe?
5. What is the outside diameter of 1" pipe?
6. What is the correct angle for lathe centers?
7. What is meant by the pitch of gears?

Cursory analysis will reveal the ambiguity or equivocal character of some of these questions. Number 1 is far too indefinite. It might require anything from an elaborate definition and description to a brief statement as to what a lathe does. Question 2 is somewhat similar in that nothing is stated as to whether information is desired regarding the shape of the lathe dog or its function. Question 3 does not indicate in what unit the answer should be given. Question 4 implies some indefinite standard as to what is meant by vital feature. Question 5 is misleading in that one does not know how exact an answer is necessary. The last two questions are somewhat more specific and definite.

**Revision of preliminary items.** All the questions were carefully reviewed in this fashion and the findings transmitted personally to the foreman who had originally submitted them. Then in conference with him revision was made with the following result for the questions just cited for illustrative purposes:

2. A lathe dog is used to TIGHTEN THE CHUCK: DRIVE THE WORK: LOCATE CENTER: CROSS FEED.
3. How many feet per minute should belts travel for the best results? 2000: 4000: 6000: 8000.
4. The most vital feature of a lathe is CARRIAGE: ALIGNMENT OF STOCKS: BACK GEARS: TOOL POSTS.
5. The approximate outside diameter of 1" pipe is 1 1/4 inches: 1 5/16 inches: 1 3/8 inches: 1 7/16 inches.
6. What is the correct angle for lathe centers? 60°: 45°: 55°: 70°.
7. Pitch of gears means SHAPE OF TEETH: WIDTH OF TEETH: NUMBER OF TEETH PER INCH: ANGLE OF GEAR TO SHAFT.

The difference between the revised and the original questions is obvious. Question 1 is dropped altogether as being entirely too indefinite. Question 2 restricts the consideration to the function of the lathe dog by providing alternative functions, the correct one of which is to be checked. Question 3 indicates the units in which the answer is desired and is so arranged that there is a wide difference between possible alternatives. Whereas, if left to his own discretion, the workman might debate between 3950 and 4000 feet, in the present form he would have little hesitation, if he knew anything about the operation, in deciding between 2000 and 4000. Question 4 becomes much more specific by precluding the possibility of any very general comment. Question 5 states the terms in which the answer is desired rather than leaving the subject to determine himself how finely he shall make his estimate. The last two questions remain intact, but have the multiple-choice answers added for the sake of uniformity. Similar revision was made of the other questions originally submitted.

After the questions had been recast into this form, they were submitted to other foremen for further suggestions and criticisms.

They were also given to a few workmen who were not to be included in the final study, in order to determine whether there were any ambiguities which had been overlooked. A few defects were discovered in this way and appropriate correction made. The result of this preliminary selection and analysis was a set of questions ready for final selection. In the present case sixty such questions similar to those in the above illustration were retained for this purpose.

The final selection of items must be made on the basis of comparison with the criterion. The questions must be given to groups of tradesmen of varying degrees of ability to determine on which questions the best workers make the highest scores. It is sometimes possible to obtain the criterion from the men's trade-union ratings. In other cases the foreman may estimate them according to the conventional classes of trade ability. In some instances it is even possible to obtain production figures, but in many of the trades the work is too complex, the individual tradesmen are performing widely varied operations, and they produce a single complex object rather than a definite number of "pieces." In the present instance the foreman's judgment alone was available. He classified the men into three groups comprising ten experts, ten journeymen, and ten apprentices. In addition, the questions were given to ten individuals outside the industry altogether who might be considered as novices.

The questions were presented on a mimeographed blank in the usual fashion with directions and illustrations explaining the method of indicating the correct one of the four alternative answers. No particular time limit was set in taking the test. It was given to lathe operators in small groups at their convenience. The results for the first fifteen questions are shown in Table LXXI. Every check in the table means that the workman indicated at the left answered correctly the question indicated at the top—for instance, workman A, the first expert, answered correctly questions 1, 3, 4, 5, 7, 9, 10, 11, 13, and 14.

The value of any individual question depends on the extent to which it is answered more frequently by the experts than by the journeymen, by the journeymen than by the apprentices, etc. A





glance at the table indicates that the questions differ in their ability to separate good from poor operators. Question 1, for instance, is answered by eight of the experts, six of the journeymen, four apprentices, and by only two novices. This question is thus somewhat differential of trade ability. In question 2, on the other hand, all classes do about equally well, and this question would not be valuable for the present purpose. While it might be possible to evaluate the different questions by inspection in this fashion if a larger number of individuals were involved, this scheme does not prove feasible. It is better practice to compute for each question the per cent of experts who answer it, the per cent of journeymen, the per cent of apprentices, and the per cent of novices. For instance, question 1 is answered by 80 per cent of the experts, 60 per cent of the journeymen, 40 per cent of the apprentices, and 20 per cent of the novices, whereas the corresponding figures for question 2 are 40 per cent, 60 per cent, 60 per cent, and 60 per cent. This may be made still clearer by putting the results in graphic form. This is done in Figure 9. Each little diagram gives the results for one question. The four classes of trade ability are laid off along the base line and the per cents of each class answering the question indicated by a mark above that point. In the diagram for question 1, for instance, above the letter N, which stands for novices, a mark is placed at a distance equal to 20 per cent on the scale; above the letter A the mark is placed at a distance proportional to 40 per cent to indicate that 40 per cent of the apprentices answered the question. Similar marks are placed for the journeymen and experts. If these marks are connected by a heavy line, this indicates the general trend for the test question.

An ideal question would presumably give a result somewhat like that of the first curve. In this case, as we go from novices through the other classes to expert, there is a steady increase in the proportion who answer the question. This ideal curve is rarely attained in actual practice, but any question approaching it within reasonable limits may be considered satisfactory. A glance at the other curves in the figure shows that some questions are manifestly worthless and that some do give the desired differentiation. There are various types of differentiation. Question 1 approximates the

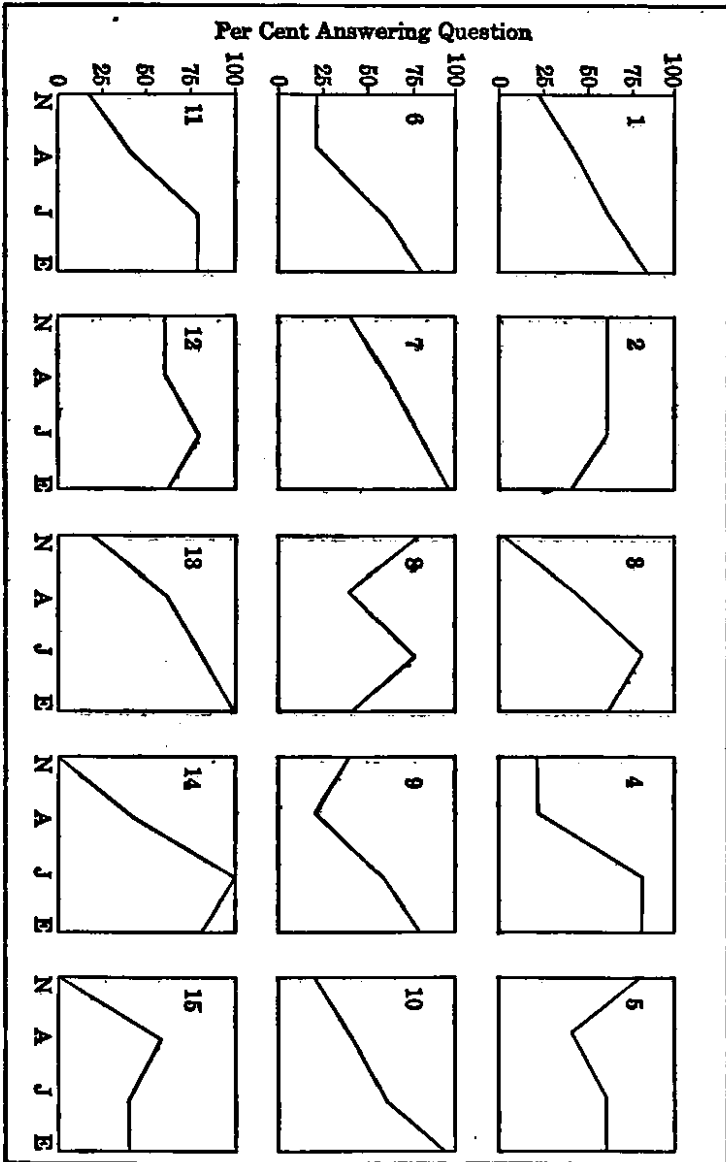


FIG. 9. SELECTION OF TRADE TEST ITEMS

ideal. Question 4 differentiates rather sharply the experts and journeymen from the apprentices and novices, although it does not differentiate sharply between the experts and journeymen nor between the apprentices and novices. Question 3 separates the novices from the other three classes without indicating a consistent difference between these three. After surveying charts such as those in the figure for each of the sixty questions, it was possible to select a limited number which seemed rather differential of the trade ability. Of those appearing in Figure 9 the following were retained: 1, 3, 4, 6, 7, 9, 10, 11, 13, and 14. These with thirty others constituted the final set of forty questions which comprised the trade test. This, then, completed the final selection of items.

**Calibration of final set of items.** One step remains, namely, to calibrate this final set of questions. Supposing a prospective employee has been given these forty questions and makes a certain score, it becomes necessary to interpret this score with a view to ascertaining his presumable trade status. We wish to know what degree of trade proficiency to expect from a person who scores ten points or fifteen points or twenty points. The procedure of determining a critical score for a trade test is analogous to that discussed previously. It is desired to obtain some score above which there is a strong probability of the individual's being an expert and below which the chances are that he is a journeyman. It is also desirable in similar fashion to draw the line between journeyman and apprentice and between apprentice and novice.

TABLE LXXII. TOTAL SCORES IN TRADE TEST QUESTIONS

EXPERTS	JOURNEYMEN	APPRENTICES	NOVICES
A 25	K 16	U 16	EE 10
B 32	L 16	V 14	FF 3
C 32	M 18	W 15	GG 4
D 34	N 18	X 7	HH 6
E 29	O 26	Y 10	II 6
F 30	P 24	Z 8	JJ 9
G 39	Q 22	AA 12	KK 8
H 37	R 18	BB 12	LL 6
I 37	S 22	CC 10	MM 8
J 38	T 14	DD 12	NN 7

Continuing the present illustration, the final scores for the subjects in the forty questions are given in Table LXXII. The entries in the table simply indicate the number of questions out of the forty which were satisfactorily answered by each individual. The first expert A, for instance, answered twenty-five questions. It is obvious from the figures that the experts make higher scores than the journeymen, and so on down the line. When, however, there is a considerable number of individuals involved in each class, it becomes difficult to draw the line between the classes by inspecting the data. In this case it is simpler to make a graphic representation of it, as is done in Figure 10. The possible points of test score are laid off along the base line. The figure is divided into four sections, one above the other. The topmost represents the experts, the next the journeyman, and so on as indicated by the letters at the left. Each square represents one man and is located in the proper trade class and directly above his score on the base line. It is obvious that the squares representing experts appear farther to the right than those representing novices. The problem now is to draw a vertical line which will make the best division between the journeymen and the experts. If, for instance, this line is drawn between 24 and 25, all the experts will be to the right of the line and all but one of the journeymen to the left. If it is drawn between 26 and 27, all of the journeymen and only one expert fall below this point. Either of these critical scores seems satisfactory inasmuch as only one man is displaced. We can then say that, if a man scores 27 or more points, the likelihood is that he is in the expert class. In similar fashion a line between 15 and 16 will clearly separate the apprentices from the journeymen with the minimum overlapping. With reference to the novices and apprentices the separation is a little less sharp, but the best point seems to be between 9 and 10. In such calibration procedure it is not usually possible to make an abrupt separation between the two classes, but the experimenter will have to use good judgment in determining the best place to draw the line. The essential point is that by drawing the line at the proper point most of those to the right will be in a trade class superior to those at the left. The points at which these lines are drawn constitute, then, the critical

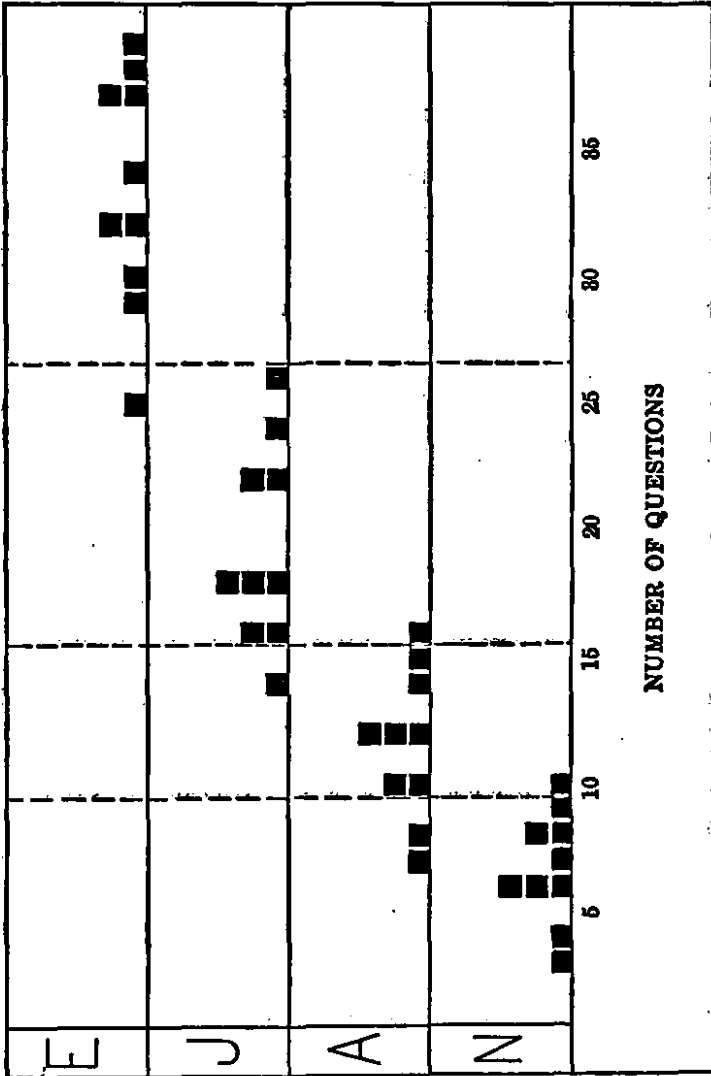


FIG. 10. CALIBRATION OF TRADE TEST ITEMS

scores and are the final figures which are desired in order to interpret the results of any applicant who subsequently may take the trade test. In the present illustration they might be stated in the following form for convenient reference:

27 to 40	Expert
16 to 26	Journeyman
10 to 15	Apprentice
0 to 9	Novice

The foregoing procedure is typical of the methods of developing and standardizing trade tests. In most instances where such tests have been developed, methods analogous to this have been utilized. Practically all such tests are made up of a considerable number of items and the technique consists of evaluating these items separately with a view to determining which ones are differential of the various degrees of trade ability. After the set of differential items is selected, it remains to standardize the total score on those items with a view to drawing a line between the different degrees of trade ability with as little overlapping as possible. These critical scores may be used in interpreting the results of applicants who are tested.

Possible refinements of method. The graphic method of interpreting trade test results is admittedly crude. It has been pointed out that curves like those in Figure 9 are misleading in one respect. The real problem is that of determining the probability as to whether a man passing a question is an apprentice rather than a novice or a journeyman rather than an apprentice. If the curve is a straight line approximately like the first one in Figure 9, it can be shown that the odds that a man passing that question is an apprentice rather than a novice are greater than the odds that he is a journeyman rather than an apprentice. For instance, suppose that the per cents of novices, apprentices, journeymen, and experts answering a question are respectively 4, 16, 28, and 40. These when plotted will all fall on a straight line. However, the ratio of the per cent for apprentices to that for novices is 4:1, the ratio of journeymen to apprentices is 1.75:1, and that for experts to journeymen 1.43:1. In other words, the differentiation between the experts and jour-

neymen is not so good as that between the journeymen and apprentices, and so on. Actually to produce this result the curve would have to be somewhat concave in an upward direction. |

In the light of this fact other possibilities have been suggested for improving the differentiation between the trade classes. (305.) It would be possible to derive items or questions that were graded into fixed levels of difficulty something analogous to the age groups used in Binet tests. It would also be possible to arrange it so that the questions in one level had rather low correlations with the questions in the other levels. This would tend to make the differential value of the various classes of questions rather sharp and would presumably make for more valid prediction.

As far as the writer knows this suggestion has not been carried out. In the practical situation the trade test is generally used for a coarser problem than is the test of innate capacity. Interest is not so much in making a fine prediction in terms of probability as in getting a general notion as to the individual's trade status and particularly as to whether he is at one or the other extreme. For this coarser purpose the present methods prove fairly satisfactory. If further refinement is needed, it is possible to follow suggestions similar to the above for selecting the separate questions. It is also possible to obtain a more refined criterion and to correlate the total score on the questions with this criterion. It would then be feasible to predict an applicant's ultimate status in terms of probability in the manner described in Chapter VIII.

#### EXAMPLES OF TRADE TESTS

It now remains to illustrate the different varieties of trade tests. The foregoing discussion dealt only with the written type of test, but the oral, the picture, and to some extent the performance types are generally similar in their method of development and calibration. However, inasmuch as their content varies somewhat, examples of each will be given.

**Oral trade test.** It is not necessary in the present connection to give a complete set of questions for any particular trade test. Complete forms of many such tests are available elsewhere. (119, 624.) In what follows only a few items for a given test will be included by

way of illustration. Each question is given in the form in which it is asked verbatim followed by the correct answer. In a few instances there are two or more answers allowable.

*Painter*

1. What do you do to knots and sappy places before painting?  
Shellac.
2. When is puttying done on new woodwork?  
After priming (first coat).
3. What is the brightest yellow used?  
Chrome.
4. What do you use to bleach an exposed oak door before refinishing?  
Oxalic acid.
5. What device is used for working just outside of a single window on a high building?  
Jack.

etc.

*Cabinetmaker*

1. What is used to close the pores of open-grained wood before finishing?  
Filler.
2. With what kind of a joint is a table leg fastened to the rail of the table?  
Mortise (tenon).  
Dowel.
3. How is an oak log sawed to get the best effect of the grain?  
Quartered.
4. What is fastened across the width of the board to keep it from warping?  
Batten (cleat).
5. How is veneer  $\frac{1}{4}$ " thick treated before gluing?  
Heated (steamed).

etc.

*Automobile mechanic*

1. What joint is there between the differential and transmission?  
Universal.



2. What are distributor brush-holder covers made of?  
Rubber (hard rubber).  
Fiber.  
Bakelite.
3. What regulates the height of gasoline in the carburetor?  
Float.  
Float valve.
4. What two metals are cam-shaft bearings usually made of?  
Bronze (brass).  
Babbit (white metal).
5. What is the best material to use to show the high point when scraping a bearing?  
Blue (Prussian blue).  
Lampblack.

etc.

#### *Bricklayer*

1. What is half of a brick called?  
Bat.
2. What is used in the middle of a long wall to keep the line level?  
Twig (twigger) (twigging) (tingle).
3. What is a brick called when set on end?  
Soldier.
4. What is a bond called when a header and stretcher are laid in the same course?  
Flemish.
5. What is the course called from which an arch starts?  
Spring (springer) (springing course).  
Skew-back.

etc.

#### *Cook*

1. What is added to milk to keep it from curdling when making creamed tomato soup?  
Soda.
2. What do you put on fried sweet potatoes to make them brown?  
Sugar.
3. What is put in buckwheat cakes to make them rise when the cakes are mixed with sweet milk?  
Baking powder.

4. What do you put in soup stock to make it clear?  
Egg (eggshell).
5. How long do you boil American macaroni?  
15 to 30 minutes.

etc.

The foregoing are illustrations of typical oral trade test questions. These questions were all the result of the method of selection described previously. They had been given to persons of known trade ability and the percentage of each trade group answering each question noted. There were involved in a given case anywhere from twelve to twenty such questions.

**Picture trade tests.** The method of developing the picture trade test is essentially similar to that of the oral. Various pictures and questions based thereon are selected and tried out to determine whether the skilled tradesmen on the average answer them more satisfactorily than do the unskilled. A few typical items from a number of picture trade tests will be described.

#### *Carpenter*

The test includes a series of pictures of tools, the question for each one being "What do you call that?" Pictures are included of such things as a jack plane, spoke shave, saw clamp, draw knife, ripping chisel, scraper, and miter. There is also a picture of a flight of steps with letters indicating the rise, the tread, and the nosing, and the applicant is asked to name the different parts that are lettered. There is a picture of a roof with the valley and the ridge indicated. The applicant must name these.

#### *Storage battery electrician*

Pictures of four battery units are shown connected in different ways — in series, in parallel, and in a combination of series and parallel. The applicant is asked to tell how many volts will be obtained from the systems under these conditions. There are also pictures of plates from various kinds of batteries with questions as to what type of battery uses that kind of plate. There are pictures of damaged plates with questions as to what might have

caused that particular kind of damage. A charging system is illustrated and the applicant required to point out the fuse and the resistance switch and to state what kind of current would be used in the circuit.

### *Machinist*

A test for machinists involves pictures of different kinds of chucks—4-jaw, 3-jaw, and drill—which the applicant must name. There is a picture of a turret lathe with a question as to what kind of lathe it is. A vernier scale is set at a certain figure which the applicant is required to read. He also has to name from the pictures various types of cutting tools and a number of different kinds of gauges.

**Written trade test.** The method of developing the written trade test was outlined in the discussion of methods (*supra*). Each item of information regarding the work involves a question and several alternative answers, the correct one of which is to be checked. Similar items for a few other trades will be given.

### *Bricklayer*

1. Half of a brick is called: **CHUNK: BLOCK: HEEL: BAT.**
2. Fire bricks are laid in: **CONCRETE: CEMENT: FIRE-CLAY: MORTAR.**
3. The top course of stone on a wall is called: **COPING: BOND-STONE: CLIPCOURSE: CAPSTONE.**
4. Before plumbing up a corner you should lay: **THREE COURSES: SIX COURSES: NINE COURSES: TWELVE COURSES.**
5. A fire stop around a flue is formed by a **COPING: SKEW-BACK: CORBEL: INDENT.**
6. To keep the line level in the middle of a long wall you use: **LEVEL: PLUMB-LINE: SQUARE: TRIGGER.**

In this instance it was possible to develop about sixty questions of this general type such that critical scores might be established at 30 and at 42 to differentiate the novice, the apprentice, and the skilled bricklayers. It was not feasible to make sharp differentiations between journeymen and experts.

*Time clerk*

A written test for time clerks involves items such as the following: Two sheets of numbers are to be added quickly — numbers such as  $8\frac{1}{2}$ ,  $8\frac{1}{4}$ ,  $11\frac{3}{4}$  — the type ordinarily added by a time clerk in computing hours and fractions thereof. There are likewise two sheets for subtracting times such as the time between 7.30 and 11.15 A.M. — another type of computation performed repeatedly by time clerks. This test was given to twenty-four time clerks early in their career and indicated rather effectively their status several months later. Of those who were in the best fourth in the test two became head clerks, two were excellent time clerks, and one was good, whereas of those in the lowest fourth all were dismissed or transferred within a short time. (112.)

*Student engineer*

A trade test was devised by one of the electrical concerns for selecting student engineers. This is essentially an information test dealing with data which such students should have learned before applying for such a position. There are three types of items intermixed throughout the test. The first type involves lists of things all but one of which belong to the same general class and the odd one is to be underlined as in the following:

Silver; copper; glass; aluminum; gold.  
81; 63; 49; 64; 16.

Another type of item involves statements which are either true or false and are to be marked accordingly:

Laminated armature cores are used because they retain magnetism better. True... False...

Resistance equivalent to a number of resistances in parallel is equal to the sum of the reciprocals of the separate resistances. True... False...

A third type of item involves problems of computation like the following:

What direct current of 110 volts will give the same horse power as a direct current of 5 amperes at 220 volts? Answer...

- Given circuits of 4 and 6 ohms in parallel and in series with a circuit of 7.6 ohms, what current will be sent through by 120 volts? Answer . . .

### *Stenographer*

The majority of the trade tests devised for stenographers and typists have consisted of a standard piece of dictation to be read at a given rate or a standard printed copy to be typewritten. In some instances where efforts have been made to devise trade tests for stenographers who are to have a little wider sphere of activity, other items are often included. The following are typical. The first group of items comprises pairs of words each pair involving two spellings of the same word such as, separate — seperate, believe — beleive. The subject checks the correct word in each instance. Another set of items involves a series of very short letters that are to be classified according to the kind of letter and the method of payment. The subject writes after each the two proper symbols from the key at the top of the blank as follows:

KIND OF LETTER	METHOD OF PAYMENT OR SHIPMENT
A Inquiry	V Prepaid
B Order	W Charged
C Complaint	X C.O.D.
D Reply to inquiry	Y Express
E Reply to order	Z Parcel post

"Please send me by parcel post 4 gross of #3 standard pencils and charge to my account."

"The fan you shipped via American Express was received in damaged condition."

The third part of the test involves letters containing mistakes in spelling and in grammar which must be checked. Other portions call for copying on the typewriter unfamiliar material such as:

seqai hjbzt ekupldr qayc umwx6

One other trade test for stenographers may be cited. (239.) The first sheet of items has words printed in shorthand with each followed by several alternatives in ordinary type. The subject checks the word which the shorthand symbol represents. There is

also a long list of common words and phrases in shorthand so arranged that each item calls for an answer to be written in shorthand. In addition to this the test proper includes standard dictation.

### *Salesman*

A test devised for salesmen should perhaps be mentioned in this connection. It consists of objections to purchase. The salesman is to assume that he is selling to the retail trade an up-to-date line for a reputable concern and that he is face to face with a prospect. He is then to indicate rather briefly what answer he would make to objections such as the following:

I cannot see you to-day.  
 I will wait until I have had a call for it.  
 Customers do not like new-fangled things.  
 You claim too much.  
 It is too high grade and expensive for me to handle.

### *Railway postal clerk*

The Civil Service uses something like a trade test in connection with hiring railway postal clerks. The blank is provided with a number of squares each containing the names of three cities and designated by a number above the square somewhat as follows:

1	2	3
Seattle	Minneapolis	Milwaukee
Portland	St. Paul	Chicago
Spokane	Des Moines	Grand Rapids

These squares are then followed by a list of cities. Each city has a short dotted line after it and the subject writes thereon the number of the square to which it belongs. These squares are supposed to represent mail sacks and the applicant is to designate in which sack the given piece of mail should go. In the actual test there are, of course, many more sacks than are indicated here. Another portion of the same test involves a series of seven sacks each with the name of an unfamiliar town. The applicant is allowed six minutes to memorize these in order to associate the number with the town. The following list is typical:

1	2	3	4	5	6	7
Hughes	Worcester	Ravina	Fulton	Jopla	Athens	Eureka

Just before taking the test the subject is told further that after September mail for Worcester is sent to Athens. He is then given another blank containing these cities repeated many times in random order and is not allowed to look at the key while he fills in the appropriate number for each. Another portion of the test utilizes a set of time-tables and a rough map of railway routes. The main line is approximately straight with branch routes leaving the various stations to reach the outlying stations. The time-tables indicate that it is possible to reach these outlying stations in various ways. The subject is told that he is on some specific main-line train and is required to figure out where it would be best to transfer mail to a branch train in order to reach a certain outlying station as quickly as possible and is also told to indicate when it would arrive there. Another part of the same test involves recognizing addresses written in script. Below each are printed several alternatives. The subject checks the alternative which he thinks the illegible script represents.

#### *Automobile driver*

One portion of the test involves information about traffic rules and other matters conducive to safety.

1. If while driving you hear the gong of the fire department behind you, you should:
  - .... Drive faster in order to keep out of the way.
  - .... Drive more slowly to let the truck pass.
  - .... Drive immediately to the curb and stop.
  - .... Stop in the street as soon as you hear the gong.
2. The chief reason why you should avoid changing gears while crossing a railroad track is:
  - .... The tracks are rough and the bumping hard on transmission.
  - .... You need all your attention to "stop, look, and listen."
  - .... Changing gears is liable to stall the engine.
  - .... You may get nervous and strip the differential.
3. Assume you are going to descend a steep slippery hill. Check three of the following things that you should do:

- .... Leave the car in gear with the engine running.
- .... Put the engine in reverse leaving the engine running.
- .... Advance the spark lever.
- .... Apply the foot brakes as necessary.
- .... Put the engine in neutral.
- .... Give the motor just enough gas to keep it running.

Another portion of the test involves recognition of dangerous situations. Pictures are shown on the blank and the subject required to write what aspect of the scene is dangerous. The pictures include parking beside a hydrant or on a curve, or double, passing a machine while ascending a hill and near the top, passing a stationary street car, traveling on the left side of a curve. The test also includes actual performance somewhat after the fashion of that for truck drivers (*infra*).

Performance trade tests as above suggested are based on quite a different principle from most of the tests hitherto described. We have been dealing thus far, except for a few of the written tests, with the principle that if a man has worked at a trade for some time he will have picked up considerable information about it. The performance tests to be described, however, deal with his actual ability to perform some operations rather than with his information.

Whatever items of score are selected, the procedure is similar to that in the development of the information type of trade test. A preliminary set of tasks, tools, and material is gathered and items of score devised. This tentative series is given to a few persons and then revised in the light of the preliminary try-out. When the final set of items is selected, it is then possible to determine the critical scores in the same fashion as previously described. A little more ingenuity is often required, as, for instance, in selecting the aspects of the product to measure and score objectively. Care is also necessary in having supplies and equipment available and having tools in good condition in order that everything may be standard. A few performance trade tests will now be described.

#### *Pattern-maker*

The applicant is provided with a standard set of tools and stock and with a blueprint. He is directed to "make a pattern for this



cast steel bracket according to this drawing." He is required first to read all the legends and measurements and point out each as he reads it in order to insure that everything is clear to him. He is then allowed to make the pattern. The time is taken and the finished product then scored in the following standard fashion. A photograph of a finished product has various dimensions indicated with their allowable margin of error. The applicant's product is then measured with reference to these dimensions to see how closely it conforms to specifications. If his product falls outside any margin of error, a defect is scored against him. One dimension, for instance, must be between  $5 \frac{1}{32}$ " and  $5 \frac{3}{32}$ ", another between 4" and  $4 \frac{1}{16}$ ". A dimension falling outside these limits constitutes a defect. There are various other penalties, such as getting the grain of one piece of wood in the wrong direction, drilling the hole all the way through when it should go only part way, totaling twenty-four possible defects in the finished product. A candidate is rated as a journeyman whose product has one of these defects and completes the work in between 71 and 120 minutes. A candidate is rated as a novice whose product does not consist of three or four blocks.

#### *Interior wireman*

The applicant is provided with two joists and cross-pieces fastened together to resemble a portion of a ceiling. He also has certain insulating tubes, knobs, wire, tape, and various other tools. He is then instructed as follows: "This is a part of a ceiling, joists, and cross-pieces. Run two feed wires across and through both joists using holes already drilled. From these main lines tap off leads in parallel and drop a lamp cord from this support. Use any material necessary, but do not use any more than you have to." The applicant is required to repeat his instructions in order to insure that he actually knows what is required. He is then left to his own devices. The finished product is scored according to a standard scheme. Certain aspects of the work are given one point credit if done in one way and no credit if done in another way. For instance, if the wires are drawn through the two outside holes 5" apart through both joists, the applicant receives a credit of one,

while if they are drawn through holes less than 5" apart he receives no credit. If he leaves a rubber tape or an open wire exposed, he receives no credit, but if friction tape entirely covers the rubber tape and all open wires are covered, he receives one point credit. He is given one point if the main lines are soldered tightly, but if they are loosely soldered no credit. In this way there are twelve possible items of score. An applicant is rated as a journeyman if he makes at least nine points and takes less than thirty minutes. He is an apprentice if he makes between two and eight points and takes more than thirty minutes. Less than two points indicates a novice.

#### *Truck drivers*

The two foregoing instances are typical of the product-time test in which the subject performs a given test, the finished product of which can be scored. One illustration will be given of a process test in which the subject is required to drive a truck through certain maneuvers. The examiner sits on the front seat beside the subject scoring him on certain aspects of driving during the test. After certain preliminary manipulations of levers, of driving forward and backing in the open, the subject enters a course nine feet wide marked off by stakes every five feet. The first portion of this course is in the shape of a letter "S" and the subject drives through at the speed he "thinks best." He is scored in this part of the test on the following errors: racing the engine when starting or shifting, abrupt start, grinding the gears when shifting, going through the S-shaped road in first speed or knocking down a stake. At the end of this course he drives his hood between two posts that are rather close together. He is penalized if he knocks down these posts. He is then compelled to back through a semi-circular road without knocking down any stakes, an error being scored against him if he makes more than one direct backing in order to enter the half-circle or if he knocks down more than one stake. He then has to back the rear of the truck up to a small platform and further errors are scored if he hits the platform or approaches it at an angle or too far to one side. He then goes to another part of the course where he is required to turn around on a side hill. Possible errors in-

clude letting the truck roll downhill more than a foot, driving with the emergency brake on, making more than one backing in order to turn around, or stalling the engine. After the subject has completed this course and all the errors have been noted, it is possible to rate him. An expert makes three errors or less, a journeyman from four to nine, an apprentice from ten to fifteen, and a novice sixteen or more.

#### GENERAL PRECAUTIONS

Reliability and validity should be considered in a trade test just as in a test of capacity. It is possible after the final set of items has been selected to divide it arbitrarily into two equal parts and to determine whether the subjects make approximately the same score in the two parts. It is less satisfactory with this type of test to give it twice and compare initial and subsequent scores, for the reason that memory for items in the first instance will influence the second repetition. Many subjects after the first test will look up or inquire about certain answers which they did not know and hence in a second test do much better. Those who perchance have not made such inquiry will be at a disadvantage. Moreover, in many of the trade tests thus far developed there are so few items that if one half is compared with the other half there is opportunity for considerable error due to the small number of items. In the army one of the main functions of the trade test was to discover those with very little proficiency in their alleged trade, and for this purpose a rather brief set of questions was sufficient to reveal the tendency.

The validity of the trade test is largely revealed in the selection of questions and the calibration procedure above described. It is obviously impractical to correlate scores with the criterion when the latter consists merely of the four degrees of trade ability. Hence it is impossible to state the validity in quantitative form. In the graphic method of calibration, however, it can be seen whether each item and also the total of the items makes possible a fair separation between the different degrees of trade ability. If this is possible with little overlapping, the test may be considered to have fairly high validity. It is, of course, also possible to give a

trade test and then compare scores with success in the trade at some later time. If those who made high scores are doing successful work and some of them have perhaps been advanced to positions of a more responsible or supervisory character while those with low scores are ineffective or have perhaps been dismissed, this is a further check on the validity. In the instance of time clerks cited above this sort of validation was made.

**Recalibration in new situation.** Just as with the various tests and measures previously discussed it is erroneous to assume that because a trade test worked in one particular situation, it will be of value in any remotely similar situation. It is desirable to recalibrate it in the place where it is going to be used. While any given trade has a good many fundamental facts and operations that will be involved wherever it is plied, there are many differences between organizations. A trade learned in one plant may differ in many essential respects from that same trade in another. For instance, the first plant may have archaic machinery while the second has modern equipment. The tradesman who has worked and is skillful in the first may be at a loss in the second, and, while such a man would make a high score in a trade test devised in the first plant, that same test would be unfair to workers in the second instance who were to deal with a different kind of machinery. The same error would be introduced if one concern used patented or other unusual machinery. Consequently, the only safe procedure is to evaluate a proposed trade test in the place where it is to be used. It may be desirable to start from the beginning, devise questions, revise them, and finally select and calibrate a set. Or it may be possible to take a set already developed elsewhere and see how valid it is in the new situation. In either instance a little research is necessary before the trade test can be made a valuable part of the employment program.

#### SUMMARY

Trade tests are designed to measure the ability possessed by a prospective employee at the time of application rather than any innate capacity that will enable him to achieve success after adequate training. They are not prognostic, but measure present

ability. They are needed in cases where it is unwise to take the applicant's word as to his trade experience and status. It is desirable so to devise the test that it can be administered by an examiner with no trade knowledge and so that it will yield an unequivocal and objective score that is quite independent of the judgment or knowledge of the person evaluating the results.

Trade tests are based on one of two general principles. It is possible to ascertain something regarding a person's trade status by giving him a standard sample of the work to perform. It is also possible to obtain indirect indications by testing his information regarding details of the trade on the theory that an experienced tradesman will have incidentally picked up considerable information about his trade and will be familiar with tools, material, and processes so that he can answer questions about them.

There are four common types of trade tests. In the oral type the questions are asked verbally and the subject's replies noted by the examiner. In the picture method the applicant is questioned regarding details in pictures of implements or machinery used in the trade. The content of the written test is similar to that of the oral test, but the form is such that the subject has merely to select the correct one from a group of alternative answers. The written test is usually adapted to group administration. In the performance test the subject does some typical standardized operation, perhaps on a small scale. This may be scored according to the process he uses, the finished product, the time consumed, or a combination of all of these. The oral and picture tests have the advantages accruing to other individual tests, namely, that they minimize opportunities for misunderstanding and consequent erroneous results and that they make possible a certain amount of clinical observation. The picture test has the additional advantage of being more concrete and making greater appeal to the applicant, but it has the disadvantage that the picture may represent a machine of a different model from that with which the applicant is familiar and thus throw him completely off the track. The written method in its group form produces great time-saving and has answers that are unequivocal and which can be easily and quickly scored by any one. The performance method comes perhaps closer to the practical

situation because it tests actual skill. It must usually, however, be given individually which requires considerable outlay in the way of equipment and materials.

The method of developing and standardizing a trade test differs from that for innate capacity tests. Whereas in the latter we are dealing with a number of tests, each composed of many items, and are comparing with the criterion the total number of items completed in a given test, in the case of the trade test we have all items of approximately the same sort and compare them individually with the criterion to determine which are the most differential items. The criterion that is most frequently used is a division of the subjects into novices, apprentices, journeymen, and experts, using these terms in the conventional trade sense. It is then necessary by consulting technical sources and conferring with foremen to devise a preliminary set of items of information or performance. These will presumably have to be revised in order to clear up ambiguities. It is well to confer with foremen in this revision and also to give the items to a small group of workers in order to locate any misunderstandings. This preliminary set of revised items is then given to tradesmen in the four groups above mentioned. For each group is determined the per cent of the members who answer a given item. If the per cent of the apprentices is higher than that of the novices, if the apprentices in turn are exceeded by the journeymen, and if the experts have the largest per cent of all answering the question, this particular item may be considered differential of trade ability. This determination may be facilitated by plotting a curve for these four percentages. A similar procedure is carried through for each question or item. It is then possible by inspection of the graphs to determine the most differential questions. These will be embodied in the final form of the trade test. It then remains to calibrate this final set of questions in order to set critical scores. This may be done graphically by plotting the total score of each individual keeping in separate blocks of the chart the different trade classes and then drawing by inspection a line between the classes with the least possible overlapping.

Tests of the four types above mentioned have been developed along these lines. The army trade tests were the first extensive

development of this sort and detailed examples are given in various reports of the army work.

It is desirable to investigate the reliability and validity of a trade test where this is possible. Half of the items may be correlated with the other half to determine reliability, although this is often not feasible because of the small number of items used. The validity is largely revealed in the calibration procedure, but it may be possible to compare scores with subsequent success in the work.

If a trade test has been developed in one situation, it is not safe to employ it in another similar one without recalibration. It often develops that methods of doing the work or the type of machinery used in one plant differ sufficiently from those in another so that a test devised in the former will be unsatisfactory in the latter. Whether or not this is true can be ascertained by a repetition of the calibration procedure to determine if the critical scores hold in the new situation.

## CHAPTER XV

### JOB ANALYSIS

#### NATURE OF JOB ANALYSIS

Job analysis has been defined as "a process of dissecting a job and describing its component elements" (290) or as a "scientific study and statement of all the facts about a job which reveal its content and the modifying factors which surround it." (582.) It comprises a consideration of the employer's contribution in the way of tools, material, pay, or general work situation, and of the workman's contribution in the way of skill, intellectual capacity, previous experience, or personal qualities. Job analysis is closely related to job specification or occupational description. The analysis is the means and the specification the end. After a detailed analysis has been conducted, the result is a series of specifications which may be used for various practical purposes. The analysis studies and ascertains the nature of the job, while the specifications reorganize this material into usable form.

#### PURPOSE

Job analysis is conducted for four main purposes. (Cf. 370.) The first of these is the improvement of methods of work. If it is desired to determine the most efficient way of doing a job, this may be facilitated by stating in standard quantitative form the different parts of the operation. We may wish to know, for instance, the time required to turn a taper, the distance one must reach for a wrench, or the time spent by a salesman in making out his reports and in other routine work. With this information in hand it may be possible to improve efficiency by eliminating wasted effort or by making technical adjustments.

A second purpose of job analysis is concerned with the health or safety of the employees. To this end study is made of the various conditions such as ventilation or illumination or the proximity of dangerous machinery to various parts of the worker's body. The



aim of this type of analysis is to find where readjustments are necessary in the interest of safety and health.

A third purpose of job analysis is concerned with more effective methods of training employees. It is often possible to organize the content of a worker's instruction more scientifically. For instance, if the difficulties of the various operations are known, it may be feasible to teach the less difficult operations first and then the more difficult ones subsequently. This plan is sometimes followed in training apprentices where the trade is divided into a number of subdivisions which are taught successively. Again, if the successful salesmen encourage the prospects to operate the adding machine themselves and ask plenty of "yes" questions, these facts may be passed along to the new men in the course of their training.

The fourth purpose of job analysis and the one with which psychology is most concerned is aimed at employment methods. From this standpoint we may analyze the work with reference to the duties, working conditions, pay, and relation to other kinds of work, and we may analyze the worker with reference to his various qualifications, innate and acquired.

#### NEED

The need for job analysis is apparent. Many occupational names are ambiguous. In personnel work in the army, for instance, wireless operators were needed in the different branches. At the outset it was assumed that any wireless operator would serve equally well in any branch. As a matter of fact a wireless operator in the heavy artillery had to care for apparatus and make repairs under adverse conditions. He also had to send and receive fifteen words a minute. A wireless operator in a motor mechanics division, on the other hand, supervised the testing and repair of radio units and apparatus. Or, again, a recruit who on his qualification card said that he was a pipe-cutter was assigned to a sewer job, but it subsequently developed that he had been a carver of meerschaum pipes. Similar ambiguities exist in present industrial terms, especially in view of the frequent subdivision and specialization of work in a given category. If a request is made for a machinist there may be persons available who are good at lathe work, but

poor at bench work, or who can operate a drill press without being able to do other kinds of machine work effectively. If a clerical worker is desired it is necessary to specify something more than this general term because qualifications are quite different for transcribing clerks such as time keepers, bill clerks, bookkeepers, stenographic clerks who do shorthand and typing or secretarial work, filing clerks who classify material alphabetically or according to topics, public service clerks who meet the public at a cashier's window, or machine-operating clerks whose work is confined largely to computing machines. Hence it is obviously necessary to specify in somewhat more detail the actual nature of the job and the actual qualifications desired for that job.

#### THE RÔLE OF PSYCHOLOGY IN JOB ANALYSIS

**Use of psychological categories.** Job analysis, to be sure, involves many things beside psychology. Much of the information deals with various items of industrial practice, but some of it also runs into the psychological categories especially when describing the necessary qualifications of workers. Mention is often made of the operative's innate capacity such as intelligence or attention. This sort of thing, we have seen earlier, may be approached more objectively, if desired, by the procedure of mental tests. Account may also be taken of his acquired proficiency in various lines and this may be approached by the trade test technique already described. Then again the qualifications may include certain personality traits and these may be evaluated by the rating scale procedure. In other words, the description of the worker will often run into psychological categories and may sometimes actually comprise the results of technical procedures such as have been discussed earlier in the book. A final job specification may frequently include scores on certain tests or rating scales.

**Psychological background for interviewer.** Psychological training will probably help the job analyst. The psychologist usually learns to observe people somewhat more closely than does the ordinary individual. In a clinic, for instance, considerable stress is attached to the involuntary movements a person makes, the way he goes at a task, and the fleeting evidences of emotional abnor-

mality. A person with a psychological or clinical background will probably observe whether the worker is performing his task automatically or with apparent conscious effort, whether he takes advantage of the rhythm of the operation, whether his eye necessarily follows his hand in making certain adjustments, whether a salesman dominates the prospect in the sales interview. Psychological training further helps in directing the analyst's attention to what the man does as well as to what the machine does. The casual observer is perhaps more inclined to watch the machine, whereas the psychologist will pay a considerable amount of attention to the workman. Again, this type of training makes one especially conscious of the necessity for concrete and specific descriptions. The psychologist is distinctly aware of the limitations in terminology when dealing with human traits. Finally, while the technique of weighting different variables, such as items of personal history or test scores, in order to predict validly some other variable, such as occupational efficiency, is not unique with psychology, nevertheless the psychologist is usually familiar with such technique and hence has a rather good background for research work. In the following discussion a brief account will be given of the current method of job analysis followed by a consideration of its primarily psychological aspects.

#### METHOD OF SECURING DATA

**Questionnaire or interview.** Various methods are used for securing job analysis data. One of the cruder procedures, which, however, is quite common, consists of issuing a questionnaire to workers and others in a position to evaluate a given job. This questionnaire may ask them in a general way to describe the nature of the job and the qualifications that they deem necessary on the part of the worker, or it may include a fairly exhaustive list of possible items as a guide to them in filling out the questionnaire. This procedure has little to recommend it. The average worker does not realize the importance of exactness in such a case and is prone to make his description in rather general terms without sufficiently defining them. The worker erroneously assumes, if he is using a word such as "good supervising ability," that this word means the

same thing to every one else that it does to him. Furthermore, one who has had no special training in job analysis will probably attach undue importance to some minor matters and fail to discriminate between the significance of different elements of the job. Consequently it is much better to obtain the information in a personal interview. If the analyst is face to face with the workman and questioning him verbally, he can adapt his procedure to the circumstances. If the worker is indefinite on some particular point, he can be questioned further on that point while it is still under consideration. If some particular "lead" is given which may to the worker seem insignificant, but is not necessarily so, the interviewer can follow it as far as seems desirable.

There may be other methods of securing job analysis data applicable to certain kinds of work. For instance, in the case of proof-readers photographs were made of their eye-movements with a rather elaborate technique (reflecting a beam of light from the eyeball onto a moving photographic film). During reading the eyes pause at various places along the line, jumping very quickly from one place to the next. Good proof-readers made seven pauses per line as compared with eleven pauses for the poor proof-readers. (290.) At present, however, the interview is practically the universal method for obtaining such data.

The interviewer's personal qualities. If the latter procedure is adopted, it is obvious that the interviewer is the crux of the situation. He needs certain qualifications in order to do his work successfully. Those who have dealt with these problems in a practical way suggest various lists of qualifications that a successful interviewer requires. (123.) For instance, he should have a rather high degree of intelligence, ability to analyze the situation, to be alert for leads, and ability to discriminate the important from the unimportant. There is less certainty as to whether he requires technical training in the job. He must surely have enough familiarity with the work to understand its terminology. It would be absurd for an interviewer to approach a man and be unable to talk to him in his own language. If the worker uses terms which are very familiar to himself and the interviewer repeatedly has to have these terms explained, it puts the interviewer in the position of not know-

ing his business and is conducive to a lack of confidence. However, it is doubtful if the interviewer needs the familiarity with the occupation that comes from personal experience and it is even possible for him to be too familiar. There is a danger in the latter case of his going into minutiae that are insignificant from the practical standpoint.

In addition to intelligence and knowledge of technical terminology the interviewer needs further various personal qualities. He must be patient because his work will often involve considerable delay. If he is about to interview an executive there are many interruptions that will occur, and if he is of an impatient type he will perhaps get into an attitude that is unfavorable to a successful interview. Moreover, he needs a considerable amount of tact. It is often difficult to get a man to talk about his job. Some persons are more or less jealous in that respect and apt to be reticent. If, however, the interviewer tactfully expresses interest in the person's work, he will probably be able to extract the desired information. He should also be rather persistent and firm in his manner because it is often necessary to keep the worker on the track. It may frequently be necessary to state, "That is all very interesting, but now what about this?" The interviewer further needs to be able to inspire confidence and cooperation so that the men will be interested in helping him in every way that they can. When a worker hails a man who interviewed him a few days previously and tells him that he has subsequently thought of one or two other aspects of his work that he forgot to mention on the previous occasion, it is obvious that the worker has confidence in the interviewer and is anxious to cooperate. Finally the interviewer must be a good salesman. He should be able to present the matter to all those whose cooperation he needs in such a way that they will be "sold" on the proposition and willing to invest their time and effort in it.

**Interviewer's training.** The interviewer must have some preliminary training before much value can be attached to his results. The amount varies in different situations. In a survey at one of the government air service experiment stations one day's intensive training was given the interviewers, whereas as a preliminary to an analysis of secretarial work the interviewer had about a month's

instruction. This training involves to some extent the preparation of questions and the revision of the wording of the questions in such a way as to bring out the desired information. Trial interviews are often valuable in which the person conducts a few interviews, not with the intention of obtaining valuable information, but for the purpose of getting the experience himself. The instructor can go over the results of these trial interviews with him and show him his mistakes or his good features. While preliminary training of this sort is usually given, it is not to be assumed that after such training the interviewer can work entirely independently. In a large organization where there are several interviewers they should have frequent conferences about their work with those who are directly responsible.

**Whom to interview.** If an organization is confronted with the problem of analyzing certain occupations, the next point to consider is what persons are to be interviewed. Obviously there are two possibilities — the workers who are actually doing the job that is being analyzed and the men who supervise their work. It might seem offhand that the superiors ought to know in great detail just what the men are doing and hence would be the most desirable parties to interview. As a matter of fact there are often minor aspects of the day's routine that do not reach the supervisor at all. For instance, a superintendent of a pressroom would consider that his foremen were essentially engaged in carrying out his orders and in getting the work out on time and might entirely overlook the fact that they also had to see that the presses were washed and oiled before leaving each night. This operation is an important part of their job, but in an actual interview it did not occur to the superintendent. On the other hand, the worker may not give all the information desired. It is hard for a person to take a detached point of view toward his work and describe all its details. This is particularly the case if a person has been at a job for a long time. Many operations become relatively automatic so that a person will perform them with very little attention. Consequently, as he thinks back over his work with a view to analysis these aspects which do not occupy much attention during the day's work are somewhat less apt to be recalled. Hence it would seem desirable

in conducting job analyses to get information both from the workers and their superiors, trusting that the details that are omitted by one will be supplied by the other.

It is further desirable to get a rather typical sampling of individuals to interview. In some instances it is, of course, possible to take everybody in the concern who is working at a given job as well as all the supervisors. If this is not feasible and a sampling is to be taken, it is well to insure that the sampling is typical and does not represent a special aspect of the work. In a study of secretarial workers (123) where persons in a great many different establishments were interviewed, effort was made to sample those in four different lines of work — secretaries in general business capacities, secretaries in institutions, secretaries to professional men, and secretaries in government positions. When the samples were selected in this way there was less danger that the analysis would reflect the peculiar features of one particular kind of secretarial work.

No definite rule can be laid down as to the number of persons who should be interviewed. After the procedure has reached a certain point, it will become obvious that the few individuals last interviewed have contributed nothing in addition to what has been contributed by earlier ones. Consequently, further interviewing will probably be of little value because it will yield little additional information.

**Types of questions.** The questions asked in the interview may be of several sorts. In the first place, they may be of a very general character designed merely to get a person talking about his work — such questions as what he finds the most interesting aspect of it or how he chanced to undertake that kind of work at all. These questions do not yield any specific information, but simply lead the person to discuss his work in a general way. In the second place, a more comprehensive list of questions dealing with many aspects of the work may be provided. Effort is made to obtain answers to these questions before the termination of the interview. A third method consists of the preparation of a long list of functions in which the worker checks those which he thinks are involved in his occupation.

The first of these alternatives is probably of value only for a preliminary survey in cases where the interviewer knows very little about the work to start with. More frequently he can have the duties and other aspects of the job somewhat classified before approaching the worker and ask him about specific items. In many occupations, even though the interviewer knows little about them, he can be certain that information as to education, physical requirements, working conditions, tools, responsibilities, and the like will be of value. The procedure of having the person simply check in a comprehensive list the things that are involved in his work is sometimes valuable in the final stages of the analysis after a good deal of preliminary information has been secured. This method can even be administered by mail or by simply distributing blanks which the worker fills out without consulting the interviewer at all. This procedure is sometimes used to increase the statistical data available for a given study. If, for instance, it has developed that a certain comprehensive group of duties is involved in various kinds of secretarial work, it may be desired then to ascertain whether secretaries to professional people encounter more frequently certain of these duties than do secretaries in ordinary business positions. This type of information can be rather readily obtained by some such check list without consuming a great deal of time on the part of the interviewer.

A work sheet is commonly used for securing the preliminary job-analysis data. It is generally agreed that it is undesirable to attempt to write up the occupational description during the interview itself. It is better to obtain the information on rough work sheets and then, after various persons have been interviewed, to compare the results and attempt to write a single occupational description on the basis of all the work sheets. This work sheet may comprise a very specific list of questions which are asked practically verbatim together with blank spaces for writing the answers obtained. A typical work sheet starts with the following questions:

1. In what year did you leave school?
2. What preliminary training did you have for this sort of work?
3. Have you ever thought of any school subjects which you wish you had taken with a view to increasing your efficiency at your present work?



On the other hand, the work sheet may comprise, not actual questions, but simply a list of topics each followed by appropriate space for notation. The interviewer is guided by these topics and the exact wording of his questions depends on the interview situation.

A typical work sheet (525, 141) comprises on its first page a space for the description of the work — the duties, responsibilities, tools and equipment, and working conditions. In describing the duties emphasis is placed, not on mere detail, but on a statement of the functions of the job. "Responsibilities" includes such things as custody of money or property and insuring the safety of other employees. "Tools and equipment" does not require mention of things such as hammers or shovels which any one can handle without much special instruction, but rather things involving special skill and training such as typewriters or acetylene welders. To facilitate the evaluation of working conditions a code is appended as follows:

- A. Imminent risk of life; e.g., experimental parachute jumper.
- B. Dangerous; e.g., flying work.
- C. Hazardous; e.g., aviation mechanic, ground man.
- D. Unhealthy or extremely unpleasant; e.g., doper, propeller test.
- E. Factory or shop.
- F. Office.

The notation of one of these code letters on the sheet is all that is necessary.

The next page of the blank involves a set of minimum requirements on the part of the worker. Physical qualities are coded in somewhat similar fashion to the preceding, as follows:

- A. Superlative; e.g., great strength (continuous heavy lifting), exceptional eyesight (draftsman, instrument maker).
- B. Superior; e.g., unusual strength (occasional heavy lifting); good eyesight (machinist).
- C. Better than average; e.g., better than average strength (carpenter, plumber); better than average eyesight (typist, fabric worker).
- D. Below average — average strength not needed (watchman, messenger, engineer); average vision not needed (doper, dry kiln operator, fire fighter).
- E. Slight — little strength required (office worker, draftsman); poor vision acceptable (janitor, laborer).

Another item calls for education with a space for entry somewhat after the fashion of the graphic rating scale as follows:

POST-GRADUATE WORK	COLLEGE	HIGH SCHOOL	COMMON SCHOOL
V IV III II I	G IV III II I	G IV III II I	G 8 7 6 5 4 3

The number indicates the grade of common school, high school, or college the individual finished. The interviewer finds it very simple then to check a certain place on this line after he obtains the information. In somewhat similar manner data regarding requirements of special training or experience may be recorded:

Special training — V IV III II 18 12 6 3 1 none!  
 Experience — V IV III II 18 12 6 3 1 none

In this case the arabic numbers indicate months and the Roman numerals years. With reference to technical skill the sheet provides a line comprising the four usual trade classifications thus:

Expert . . . Journeyman . . . Apprentice . . . None . . .

The presence of this item on the blank suggests, of course, the desirability in some instances of setting a critical trade test score. If the results of this job analysis are to be used in employing persons where technical skill is desirable, it will be more satisfactory, as has been shown previously, actually to give a man a trade test and determine on that basis whether he has the requisite trade ability rather than to take his word for it. The job analysis would then state the amount of technical skill necessary, and in using this analysis for employment purposes the trade test would determine whether the applicant had the requisite skill.

Further items on the blank deal with personal qualities and are arranged in the manner of the graphic rating scale:

Judgment	<hr/>				
	Unfailing; errors cause personal danger	Good; errors cause money loss	Average; errors cause confusion	None	
Creative ability	<hr/>				
	Highest; inventiveness	High; originality	Average; initiative	None	
Super- visory ability	<hr/>				
	500	100	25	10	2 None

The last of these calls for the number of men whom the individual supervises.

Each of these "requirements" has also a blank space labeled "reason." In this the interviewer must justify the entry he has just made. For example, on the work sheet for automobile mechanics the entry "keen hearing" was justified by the statement that this was necessary in order to diagnose motor trouble; common-school education was required in order to make out time slips and read written directions; a year's previous training in a garage or repair shop was requisite in order to shorten the learning period, and good judgment was listed because it was required in "shooting trouble." This procedure of making the interviewer justify each entry insures that the item listed represents a real requirement and not an imaginary one. It puts the interviewer and the one being interviewed to the necessity of really considering the value of certain items. It also clarifies the qualification itself by showing a concrete way in which it is to function.

Another page of the blank is similar to the one for minimum requirements, but deals with further requirements that are desired, but not absolutely essential. It comprises the same set of items with spaces for writing the answers and also justifying them. The interviewer can then list the qualifications of a given sort according to whether they are minimum or simply desirable.

In the administration of such a work sheet it is not customary to use a stencil or to measure the position of the check marks, as is done with the usual graphic rating scale, but simply to code them by having some letter for each of the phrases or numbers on the line. These code letters may then be entered in boxes at the top of the work sheet.

Many of the items on such work sheets are not of psychological character, but there are manifestly certain aspects in which psychology is or might well be involved. The conventional rating scale procedure is suggested by the consideration of various character traits. The question of trade qualifications immediately points to the technique of trade tests. In certain types of work it might prove that additional items regarding intelligence were desirable. It might even be possible to analyze a job with reference

to whether it required a high degree of attention or a certain amount of memory or the ability to make quick decisions or other special capacities. A person with psychological training might frequently find items of this sort which could very well be included in the analysis of the job.

OCCUPATIONAL DESCRIPTION

**Form.** After a considerable number of persons have been interviewed and a work sheet filled out for each, it is possible to note statements regarding duties and qualifications on which there is substantial agreement. The interviewer is then in a position to write up the results of his interviews in the form of a final occupational description or job specification. While the form of such description may vary with the circumstances and the preferences of those most concerned, it is rather an established practice to put the description in the form of simple declarative sentences.

**Examples.** A few typical occupational descriptions will be cited.

*Occupational description for automobile mechanic*<sup>1</sup>

<b>Duties</b>	The automobiles and trucks used by this company are kept in condition in the Garage Branch of the Maintenance Section. Under direction the automobile mechanic overhauls, repairs, and operates such standard machines as the Dodge and Cadillac touring cars and Mack, Standard B, and G. M. C. motor trucks. He tests, overhauls, and repairs motors, generators, and ignition units. He does acetylene welding, and uses tools such as lathe, reamer, and valve-reader.
<b>Hours</b>	7.45 A.M. to 11.30 A.M. 11.30 A.M. to 12.50 P.M. Lunch } Monday to Friday. 12.15 P.M. to 4.30 P.M. 7.45 A.M. to 11.45 A.M. Saturday.
<b>Minimum qualifications</b>	The automobile mechanic must have graduated from common school and in addition he must have had three years' practical experience in a garage or automotive machine shop as repairman. In lieu of one year of practical experience, six months' special training in automobile repairing or one year as machinist apprentice will be accepted. Man 18 to 50 years of age.
<b>Additional qualifications desired</b>	The automobile mechanic should be physically strong, capable of occasional heavy lifting. He should have good eyesight in order to do close work and make fine adjustments, although glasses are permitted. Keen hearing is also desired in order to enable him to test motors by sound. Accuracy is important in this work as errors may cause delay and impair work.
<b>Working conditions</b>	Garage with concrete floor. The worker is on his feet about half the time. Much of his time is in a crouching or prone position incident to repairs underneath cars. The automobile mechanic is outdoors part of the time, especially when testing machines on the road.
<b>Principal lines of promotion</b>	From: Truck driver, chauffeur, mechanic's helper. To: Garage superintendent, engine mechanic.

<sup>1</sup> After Scott and Clothier.

It is obvious that this description embodies the information discussed earlier in connection with the interview. It begins with a description of duties, stating exactly what the man does. It also gives the hours which he works. The next section gives the minimum qualifications and covers the various topics of education, experience, and the like that have been discussed before. There are also additional qualifications which are desired but not absolutely necessary. Further information appears regarding working conditions and also the principal lines of promotion. This latter gives a notion as to the most profitable positions from which to recruit personnel for the job in question and also the positions into which one may be promoted after having had adequate experience. In the original occupational description sheet just given there is also a series of boxes at the top for quick reference. These boxes refer to items such as education, experience, judgment, accuracy, supervision, physical qualities, and working conditions. In each box is entered a single letter. This letter refers in code to different degrees of the qualification or item mentioned. Such codes for physical qualities and working conditions have already been given (p. 476). The following are the notations for recording in code the remaining items.

#### *Education*

- A. Graduation from college
- B. Graduation from high school
- C. Two years' high school
- D. Graduation from common school
- E. Six years' common school
- F. None

#### *Experience*

- A. Ten years
- B. Five years
- C. Three years
- D. Two years
- E. One year
- F. None

#### *Judgment*

- A. Errors may cause loss of life
- B. Errors may cause personal injury

- C. Errors may cause money loss
- D. Errors may cause confusion — inter-departmental
- E. Errors may cause inconvenience — intra-departmental
- F. None

*Accuracy*

- A. Errors may cause loss of life
- B. Errors may cause personal injury
- C. Errors may cause money loss
- D. Errors may cause confusion — inter-departmental
- E. Errors may cause inconvenience — intra-departmental
- F. None

*Supervision*

- A. Supervising 100
- B. Supervising 50
- C. Supervising 25
- D. Supervising 10
- E. Supervising 5
- F. None

*Occupational description for designer in structural steel*<sup>1</sup>

<b>Pay</b>	The employment salary is \$30 per week.
<b>Duties and responsibilities</b>	<p>The designer in structural steel designs the steelwork necessary for <i>cooling towers</i> and <i>cooling bridges</i>; the steel framework for <i>substation</i> and <i>generating stations</i> and miscellaneous structures such as stairways and platforms. All of this work is constructed either by contract or by the company's building construction department.</p> <p>In designing steelwork the designer should be familiar with:</p> <ul style="list-style-type: none"> <li>(a) The loads to which the structures will be subjected.</li> <li>(b) Structural steel handbooks which give tables of the size of structural members, such as <i>eye beams</i>, <i>H columns</i>, <i>channels</i>, <i>angles</i>, and <i>girder beams</i>, which he will use.</li> <li>(c) The standard methods of making connections between beams and columns, beams with beams, etc.</li> <li>(d) The necessary <i>struts</i> and <i>braces</i> and the methods of connecting these with columns and beams.</li> </ul> <p>The designer in structural steel also designs the steel framework for additions to be made to existing buildings. Before making these designs he takes field measurements at points where the new work is to be added, noting carefully whether reinforcements will be necessary in any existing construction and such things as connections of floors and walls and special foundations for heavy machinery.</p> <p>Where steel stairways, platforms, or ladders are to be built inside of substations or power houses, the designer takes field measurements at the location, allowing for clearance between new and existing work.</p> <p>At times the designer designs <i>steel smoke breechings</i> for boiler rooms of power houses. In making such designs he should make allowances for expansion of the breeching due to heat, seeing that the proper clearances are allowed between these structures and existing steelwork. He should be familiar with <i>expansion joints</i> necessary for <i>smoke breechings</i>.</p> <p>He makes preliminary lay-outs of all designs; he should be familiar with standard drawing practices, structural steel designs, and standard drawing instruments. He must know how to operate a Universal drafting machine and use all miscellaneous materials used by draftsmen such as drawing paper, tracing cloth, different grades of pencils, and drawing inks.</p> <p>A man of 25 to 35 years of age is desired.</p>
<b>Personal qualities desired</b>	<p>Initiative above the average is essential, as he must in most cases use his own judgment in working out the best methods of design.</p> <p>Accuracy in determining the necessary size and kind of structural member is</p>

<sup>1</sup> After Snow.

## EMPLOYMENT PSYCHOLOGY

of prime importance, as a considerable saving of money is effected if these members are of the exact weight necessary to support a definite load plus an additional load for a factor of safety.

Neatness in making drawings that can be understood by others is an essential quality.

**Working conditions**

The work is permanent and is highly technical and mechanical in nature. Working at a drawing board making drawings may cause some eye-strain. The drafting room where the lay-outs and designs are made is well lighted, ventilated, and arranged.

There is some outdoor work attached to this position as when taking field measurements.

**Education and experience desired**

A fourth-year technical education or its equivalent is desired. One that is specialized on the theory of structural steel design is preferred.

Two years' experience designing steel buildings for power plants and substations is desired.

**Opportunities for advancement**

There is at present no direct line of promotion from this position.

There is opportunity for men of this type to secure positions with high responsibilities with structural steel corporations.

**Sources of supply**

Draftsmen of structural steel corporations are external sources of supply.

Draftsmen are an internal source of supply.

This description is in form substantially like the preceding, giving information regarding duties and personal qualities, working conditions, education, and experience, and lines of promotion or positions from which to select people to promote to this position. An example of an occupational description that is somewhat less comprehensive is the following for workers in a prison shoe factory. Each job has a brief description together with certain further qualifications. It is to be noted that mental age as based on intelligence tests is included as one of the qualifications. Amount of schooling is also specified as well as the time required to learn the job and the time required to attain skill.

Job	Description of Job	MENTAL AGE	SCHOOLING	TIME TO BREAK IN	TIME TO ATTAIN SKILL
Sole rounding machine operator	Fits insole to wood model. Adjusts machine. Knives cut leather to form and size of model.	13-15	4	1 week	1 month
Channeling machine operator	Holds insole against channeling knife which cuts shoulder on edge of insole and opens channel for welter.	13-15	4	1 week	1 month
Cutter	Cuts leather for uppers. Cuts lining for uppers.	13-15	4	3 weeks	3 months
Closer	Operates sewing machine. Joins quarters by sewing up back.	11-13	3	1 day	1 month
Bagger or top closer	Operates sewing machine. Sews cot on lining and leather quarters together.	12-14	3	2 days	1 month
Eyeleter	Operates eyelet punching machine.	12-14	3	2 days	1 month
Vamper	Operates sewing machine. Sews vamps to quarters and tongue.	13-15	3	1 week	2 months

PSYCHOLOGICAL POSSIBILITIES IN JOB ANALYSIS FOR  
EMPLOYMENT PURPOSES

The foregoing illustrates the more recent methods of job analysis and occupational description. It is evident that some of the points are distinctly psychological in character. In so far as the job specification deals with the worker, it is bound to include mental factors and psychological terminology. This is related to the interests of the psychologist in two ways. As hinted in Chapter VIII, the results of job analysis may be of value to a psychologist initiating a project of developing mental tests for some particular occupation. In such a case it is necessary to determine what mental characteristics are required by the occupation with a view to devising tests for those characteristics. If a careful job analysis has been conducted by trained interviewers and occupational descriptions are available, these may give the psychologist some notion as to the nature of the job. He will often find this a valuable starting-point for his own analysis with a view to developing tests. If the occupational description mentions keen hearing, good attention, powers of observation, or the necessity of making motions quickly, this points to rather obvious psychological test possibilities. The experimenter will doubtless supplement this sort of information with further observation of his own, but it often calls his attention to aspects of the job that he might otherwise have overlooked and affords him a good beginning for his work.

On the other hand, the psychologist has something to contribute to the job analysis program. Many of the principles discussed earlier in the book might very well enter here as a supplement to the method. If the job specification is to be the final instrument used for hiring workers, it might theoretically embody a number of these principles. The remainder of the chapter will point out a few of these which might fit into a comprehensive job analysis program.

Statistical validation of miscellaneous factors. In the first place, it may often be desirable to evaluate statistically certain miscellaneous items of personal history such as are brought out in the analysis. For instance, height and weight are sometimes noted by interviewers as desirable for a given kind of work and if the interviewer on his work sheet is attempting to justify such items he will



state that the work is "heavy." Or if the interviewer finds that an eighth-grade education is necessary, he justifies it on the grounds that the worker must read time slips. Or if he states that a married worker is preferable, he substantiates that judgment by the fact that such a worker will be more stable. Now it is statistically possible to find out whether a certain height or weight is necessary for the job, whether an eighth-grade education actually is the necessary minimum, and whether married workers are more stable. The procedure discussed in Chapter XIII on miscellaneous determinants of vocational aptitude is directly applicable here. It is necessary merely to obtain groups of workers of a given sort some of whom are reasonably successful while others are unsuccessful, and tabulate them with reference to such items as height, weight, age, or marital status, to see to what extent these items differentiate the successful from the unsuccessful group. While the judgment of the interviewer may be usually sound when dealing with matters that are fairly obvious, there is no real guarantee that his information is always well founded. Some of the persons whom he interviews may have made hasty generalizations and passed them on to their colleagues, so that there will be unanimity in some statement that is actually erroneous. The technique of statistical validation will insure against any such error.

**Rating scales.** In the second place, the technique of rating scales would seem rather generally applicable to the various personality factors that are sometimes encountered in job specifications. Some of the work sheets described above embody such of this technique as is applicable. Statements regarding judgment or creative ability are recorded by checking on a line with descriptive phrases beneath it. The only suggestion to be made at this point is a somewhat wider extension of this policy to cover other traits that might be of possible significance in many types of work. Such things as tact, leadership, and coöperativeness and many other traits discussed in Chapter XII would seem applicable. The average location of the check marks on the work sheets would indicate the degree of the trait that was requisite. In some cases the ratings might be combined quantitatively into a total rating as described previously. In many instances, however, they could be recorded

in code similar to that mentioned in the present chapter. It would thus be possible to determine something analogous to critical scores in many specific traits and embody them in the final job specification. In using these facts for promotion or transfer within the organization, those being considered for such change could be rated systematically by their superiors. In hiring employees from outside the procedure of using a rating scale during the employment interview might be followed.

**Trade tests.** In the third place, the possibility of trade tests has already been suggested. Whenever the job specification asks for technical skill and mentions apprentice or journeyman or expert, this immediately raises the question of how that trade status is to be determined. To be sure, a man's union card will often give an approximate notion of this status. There is no guarantee, however, that this will always be reliable, and it is more dubious still to take a man's own word in the matter. The technique of trade tests has reached the point where it would be applicable to almost any type of operation that requires specific trade skill. Consequently, if such tests were developed for the job in question the occupational description might well include a critical trade test score.

**Intelligence tests.** In the fourth place, intelligence tests might often well be an item in the job specification. We have previously seen that certain work, such as clerical, shows a definite correlation between intelligence and occupational efficiency. We have also seen that some occupations in the hierarchy appear to require a certain general level of intelligence and that persons too low or too high are unsuited for that type of work. In such cases it is customary to establish a critical score in intelligence as a basis for hiring. It would thus seem logical in an organization where intelligence tests had been standardized to embody in the job specification a critical score in the intelligence test.

**Special capacity tests.** Finally, the tests of special capacity which were described at considerable length in Chapters VIII and IX might play a rôle in this procedure. In any occupation where such tests have been worked out, critical scores either on separate tests or on the weighted sum of the tests might be introduced as one important item in the job specification. The aim of job analysis and specifica-

tion is, of course, to present to the applicant all the necessary information about the proposed job and its possibilities and to obtain all the necessary information about the applicant with a view to occupational prognosis. While a given organization must be governed considerably by the extent to which it can invest in the employment program, and while the validity of psychological methods depends considerably on the local situation, there are doubtless a great many instances in which it would be desirable to develop a rather complete and extensive job specification. This comprehensive specification would include items similar to those given in the various specifications cited above by way of illustration and might also comprise critical scores in rating scales, trade tests, and various other tests of innate capacity which have been statistically studied with reference to the job in question.

#### SUMMARY

Job analysis involves dissecting a job both from the standpoint of the work and from the standpoint of the worker. It leads to a detailed job specification or occupational description which may be used for improving working conditions, promoting health and safety, perfecting methods of training, and supplementing employment procedure. Only the last of these is our present concern. Job specification is a necessary part of the employment program because of the ambiguity of many occupational terms and the diversity of operations often included under a general title.

Job analysis involves much that is not psychological, but in describing the worker there is perforce a considerable use of psychological categories. Moreover, a psychological background will assist a person conducting job analysis because of his training in observing people.

The data are usually secured by means of personal interviews with employees and executives. The interviewer should be familiar with the technical terminology, but need not be experienced in the occupation in question. He needs likewise various personal qualities such as patience and tact. Preliminary training for interviewers is desirable and this may well take the form of trial interviews with criticism of the results.

It is wise to interview both workers and their superiors. The former may attach little significance to their acts that have become automatized, while the latter may overlook minor aspects of the job that would occur to them if they were actually going through it. In selecting persons to interview it is desirable to secure a sample that will be typical rather than to represent only one aspect of the job.

A set of questions may be prepared for the interview and asked verbatim, or a list of topics may be provided which the interviewer follows guiding his actual questions by the trend of the conversation, or a comprehensive list of functions may be used in which those involved in the job are to be checked. The second method is the one generally followed.

The interviewer may well be provided with a "work sheet" calling for various items of information, such as duties, responsibilities, equipment, tools, and working conditions. Minimum requirements of the worker are to be ascertained and often rated according to a code or a brief rating scale. Each item, such as physical qualities or experience or judgment, must have a reason given for the particular entry made.

With these data for many interviews it is possible to write the occupational description or job specification. It is rather common practice to put this in the form of simple declarative sentences with a brief paragraph covering each item, such as duties, hours, minimum qualifications, additional qualifications, working conditions, and lines of promotion. For convenient reference most of these facts may be reduced to a code notation and indicated in boxes at the top of the blank.

Many of the factors involved in job analysis are, of course, non-psychological in character. However, the analysis may be of assistance to the psychologist initiating a project of developing tests for a given occupation. While he may have to go farther himself in determining the mental aspects of the job for which it is advisable to develop tests, the job analysis if already conducted may well afford him a starting-point.

On the other hand, many of the psychological methods already discussed may make some contribution to job analysis. It is

possible to evaluate statistically miscellaneous items of personal history which are often included in the occupational description on the basis of the judgment of those interviewed without necessarily a scientific justification. The analysis often includes various mental traits such as are usually embodied in rating scales. If the rating scale technique is used to determine the amount of the trait necessary for the job, the applicants may be rated similarly to see whether they attain this critical amount.

Wherever the job specification calls for previous experience in a trade the desirability of a trade test is obvious. Instead of taking the applicant's word for the matter he may well be tested to determine his status. The job specification may then embody a critical trade test score.

In view of what we know regarding the correlation of intelligence with vocational aptitude and of the nature of the occupational hierarchy, it would seem logical for the specifications for certain jobs to contain critical scores in intelligence.

Finally, tests for special capacity may frequently be developed as a part of the job analysis procedure and critical scores embodied in the final specifications. Theoretically the job specification ought to contain everything that will promote the selection of workers who will be efficient and happy. In addition to the usual information regarding duties, hours, salary, and personal qualifications of the sort revealed by the applicant's own statements, it will in many cases promote this more effective selection to include critical scores, or something analogous, in rating scales, trade tests, items of personal history, and tests of intelligence and special mental capacity.

## CHAPTER XVI

### THE OUTLOOK FOR EMPLOYMENT PSYCHOLOGY

#### SUMMARY OF PSYCHOLOGICAL TECHNIQUE APPLIED TO EMPLOYMENT METHODS

We have now completed our survey of present-day psychological technique in so far as it bears on problems of employment. After clearing the ground of certain pseudo-psychology which is often presented to the business man as a remedy for his employment difficulties, we discussed the technique that is most widely used in this field, namely, mental tests. The distinction was made between tests of innate capacity and tests of acquired proficiency and the former were subdivided further into tests of special capacities such as attention or memory and tests of general capacity or intelligence. Illustrations were given of a considerable variety of such tests with which an employment psychologist would ordinarily be familiar before undertaking a research project. The technique of devising and administering tests was described in detail. Attention was called to the fundamental importance of validating the tests or other measurements by comparing them with the "criterion" — some expression of the workers' ability in the job. It is always necessary to determine whether those who are efficient in the test are efficient in the job, and *vice versa*, before the tests can be validly used for occupational prognosis.

We then discussed in more detail the criterion of occupational efficiency either in the form of ratings by the employee's superiors or of production figures. We also noted two possible types of subjects on whom to standardize the tests — employees and applicants. While the latter are perhaps better from the theoretical standpoint because they are more similar to the persons in hiring whom the tests are to be ultimately used, nevertheless as a practical matter employees have been more frequently taken as subjects in research for the simple reason that the criterion is available more quickly.

We then turned to the specific procedure of validating the mental tests by comparing test scores with the criterion. Dealing first of all with tests of special capacity, approach to the problem may be made through two avenues — reproducing the total mental situation involved in the job and analyzing the job into its mental components and measuring these separately. In the former case the score in the single complicated test is correlated with the criterion to determine its value. In the latter case each test is correlated separately with the criterion in order to retain the most valuable tests and discard the others. This final group of tests is then weighted in order to allow for any overlapping of the different tests and to combine them in such a way as to get the best possible prediction of vocational aptitude. In either instance, when we know the final correlation of our single test or our group of tests with the criterion, we are able to set a critical score as a basis for hiring or rejecting applicants. This critical score may best be determined by computing the probability that an applicant with a certain test score will reach a certain level of occupational success. The employment department then knows how big a chance it is taking with an applicant and can decide after considering all other related factors whether it wishes to take this chance.

We then considered general capacity or intelligence as related to vocational aptitude. Occupations appear to follow an intelligence hierarchy inasmuch as the average intelligence of occupational groups increases consistently from unskilled labor to the professions. It suggests that a person tends to attain an occupation as high in the scale as his intelligence warrants. This gives some notion of the intellectual requirements of various occupations. In some types of work intelligence scores correlate significantly with the criterion so that the procedure used with special capacity tests is applicable. Furthermore, in some instances the work requires, not necessarily a maximum intelligence, but rather an optimum intelligence, inasmuch as persons who are too good for their job are apt to be dissatisfied and quit.

Interest as well as ability is important in vocational prediction. Consequently, methods of measuring interest were discussed, and although the methods are still in the experimental stage a few

instances were presented in which interest data were evaluated with reference to vocational success.

We then turned to the technique for dealing with certain traits, such as industry, coöperativeness, tact, and enthusiasm, which cannot at present be measured by tests, but are nevertheless of vocational significance. For such traits the judgments of acquaintances or colleagues can be systematized by means of rating scales. In the man-to-man scale the person being rated is compared with others on a previously constructed master scale; in the method of defined groups he is located with reference to the distribution of similar workers into a series of groups of equal size possessing the trait in an increasing degree; in the graphic method his standing is indicated by a check mark somewhere along a line on which the rater is guided by descriptive adjectives. In all these instances one trait is evaluated at a time in order to abstract from errors due to general impression.

We then discussed miscellaneous factors which may be used as a supplement to mental tests or in lieu of them where tests are not feasible. Educational status and items of personal history such as often appear on the application blank may be statistically evaluated by determining which individual items are differential of occupational ability. Application letters were shown to be very unreliable, but the best procedure for dealing with them is to pool the independent judgments of several persons who evaluate them. The recommendation procedure may be improved by the use of a blank of inquiry calling for brief answers or check marks. The employment interview may well be supplemented by an interviewer's rating scale.

We then turned to the trade test which, instead of prophesying future occupational status, is designed to measure a person's trade skill or information at the present time. The technique consists of testing novices, apprentices, journeymen, and experts and finding which particular items or questions are differential of these groups. It is then possible to establish a critical score to determine in which of these trade classes an applicant belongs.

Finally, we discussed job analysis in so far as it bears on employment psychology. Many alleged requirements on the part of the



worker can be statistically evaluated before their inclusion in the final occupational description. The rating scale technique may prove valuable in dealing with certain personality factors in the analysis. Where trade experience is a necessary requirement the technique of trade tests would seem in point. In many types of work the job specification might well include critical scores in tests of intelligence or special capacity.

The foregoing are some of the psychological principles that are applicable to problems of employment. They have been the result of gradual development and of the coöperation of many psychologists all along the line from the first ones who constructed mental tests through those who perfected the statistical methods to those who have actually validated tests and other techniques in various practical fields. It now remains to consider the present status of the science and to look toward its future.

#### PRESENT TRENDS

**Individual consulting.** At present a very considerable number of individual projects in employment psychology are under way. Some psychologists are actually engaged in full-time work either as consultants or as members of the staff dealing with personnel problems for industrial concerns. Others are working in the academic field, but pursuing a certain amount of personnel research on the side. In many cases where a college or university is located in a city that possesses a considerable number of industries, it is feasible for a psychologist at the university to do some practical work in local plants. Oftentimes advanced students taking a laboratory course in industrial psychology do some of their laboratory work in the field — that is, in the local concerns.

**Coöperative research.** Another present trend is coöperative research. One type of coöperation involves simply the interchange of results and methods between psychologists. It is a rather common practice when one has completed some piece of work dealing with employment or other industrial problems to publish his results in scientific periodicals so that others may have the advantage of his experience and so that various workers will not be duplicating one another's experiments. Psychology has not reached the point,

except in a few instances, where methods developed in one concern are kept secret. Psychology is primarily interested in promoting human efficiency and happiness, and this perhaps can be most rapidly furthered by an interchange of ideas and coöperation between psychologists.

In another type of coöperative research that has been more systematically organized, a number of business concerns and scientists work together on some particular problem. For instance, the study of turnover among salesmen is naturally of concern to business men and of interest to psychologists. These latter may, however, be occupied with their own work and unable to spare sufficient time to study personally the turnover problem. In such cases it has been feasible for the business groups to contribute financially to the support of an organization to undertake this research problem. The scientists who are interested can then perhaps supervise the more detailed work carried on by a staff that is hired for the purpose. Typical of such coöperative research is the work of the Bureau of Salesmanship Research that was organized at the Carnegie Institute of Technology. (52.) The head of a large insurance firm came to the Institute with a request for courses in salesmanship which went somewhat further than the conventional type of course. His attention was called to the need for more facts, such as the difference between the successful and unsuccessful salesmen, their aptitudes and traits, a study of different kinds of appeals, and of methods of selecting men and providing incentives. As a result of this conference other firms were approached on the proposition, so that finally about thirty concerns contributed over a period of years to support a bureau. A competent staff was organized and embarked on a systematic study of salesmanship. In addition to contributing to the support of the organization the concerns opened their records and their experience to the research workers so that all available information was put into a common pool. The concerns furthermore coöperated in carrying out experiments with different groups of salesmen and different methods. The bureau was governed by a board representing both the institution and the coöperating concerns. This is not the place to recount the work of this bureau, for it is cited merely to illustrate

this type of coöperative research. It was somewhat interrupted by the War, but it developed a series of "aids" for sales managers consisting of model application blanks, model letters of reference, various improvements in the interview procedure, and batteries of tests for selecting salesmen. These "aids" were distributed to the coöperating concerns. Other similar bureaus were the outgrowth of this first one. For instance, one was organized to meet the problems of the local retailers. It prepared employment tests, trained members in specific methods of correcting difficulties, and studied sales personalities. Results in retail stores were checked by "service shopping" in which certain individuals were hired to go shopping in the stores and to take careful notes as to what transpired in each sale. This service shopping gave a quantitative expression of the per cent of dissatisfied customers and statistics showed how this percentage decreased as the result of the bureau's work.

In installing a coöperative research project of this type it is important to insure its stability. Scientists of major caliber are not inclined to enter upon such a program if it is liable to be interrupted before its completion, and a research problem cannot be solved overnight. To this end a rather long-time contract is desirable. Furthermore, a concern which is rather frequently changing management is not a good concern to participate in coöperative research. If the vice-president or sales manager is changed annually or more frequently, it is necessary to "sell" the new incumbent the entire program. In organizing such research a further problem consists of getting an adequate research personnel. It is often difficult to induce persons of the general status of graduate students to engage upon research work of this type, and frequently, after they have been at it for a short time, they accept individual openings elsewhere in personnel work. In some instances it has been possible by offering fellowships to induce a high grade of research personnel into this field. A bureau of this sort further must be on its guard against getting off the track into "service" work, such as giving public addresses, arranging conventions, or compiling statistics, which are interesting but detract interest from the main point.

## THE OUTLOOK FOR EMPLOYMENT PSYCHOLOGY 495

The National Research Council was organized under federal charter of the National Academy of Science and comprises various subdivisions, among them a Division of Anthropology and Psychology. The Council is not merely a laboratory or repository of findings, but endeavors to coordinate research and further the organization and support of undertakings which demand the cooperation of individuals or institutions or both. During the War many personnel studies were organized under the auspices of the Council as described in previous chapters. Various more recent projects are under way, such as the relation of intelligence and schooling to occupational ability, the organized search for research talent among college students, the analysis of mechanical ability, and the devising of methods for measuring it. The Council includes a Research Information Service wherein it keeps a record of the research that psychologists are doing and of their interests and activities, so that when any one becomes interested in a problem he may ascertain who else is working on the same problem and may perhaps cooperate.

**Psychological methods in the Civil Service.** The United States Civil Service has recently employed psychological methods to a considerable extent. It has devised a scale or set of tests of "general adaptability." These tests are focussed at different industrial levels in such a way as to cover the entire range and attempt to measure "the ability to learn, to solve new problems, and to meet new situations." They have been used especially in selecting office clerks. Incidentally a very considerable amount of time has been saved in correcting the results of Civil Service examinations by the use of answers of the multiple-choice form.

The Service has likewise developed various tests for special aptitude, such as that for mail distributors, in which they classify names of cities according to various boxes or have to discriminate specimens of rather illegible handwriting. Tests for policemen have likewise been developed embodying such things as ability to evaluate evidence, the significance attached to different acts, and judgment as to the action an officer should take in a particular situation. The Service is further developing examinations in various

engineering subjects especially for the selection of examiners for the Patent Office.

The Bureau of Public Personnel Administration was organized in October, 1922. Its origin was stimulated by the fact that various Civil Service Commissions and other public service agencies working independently on personnel problems often duplicated one another's efforts. Consequently this Bureau was organized to serve as a clearing-house for existing information relating to public personnel administration. The Bureau conducts further experiments and publishes the results, issuing from time to time a series of business personnel studies. This Bureau has developed such things as tests for policemen, firemen, various skilled trades, stenographers, typists, and clerical workers.

The Personnel Research Federation arose through coöperation between the National Research Council and the Engineering Foundation. Its membership comprises many agencies and institutions such as universities, business concerns, and a considerable number of private individuals. Its purpose is to "further research activities pertaining to personnel in industry, commerce, education, and government wherever such researches are conducted in the spirit and with the methods of science." One of its important contributions is the publishing of an official organ, *The Journal of Personnel Research*, through which many studies in this field are made public.

The Psychological Corporation was founded in 1921. It is incorporated, not for profit, but for "the advancement of psychology and the promotion of the useful applications of psychology." It can issue no dividend over six per cent per year. The stock is subscribed and all held by psychologists with the provision that at any time the American Psychological Association (the official national organization of psychologists) can purchase the entire stock, in this way bringing the corporation under control of this Association. All of the original directors were psychologists of note — every one of them appearing in *Who's Who* — a rather unique board of directors. The Corporation has branches in Massachusetts, Pennsylvania, Maryland, District of Columbia, Ohio, Michigan, Illinois, Iowa, Kansas, Missouri, and California. Other branches are in the

## THE OUTLOOK FOR EMPLOYMENT PSYCHOLOGY 497

process of organization. One of the main objects of the Corporation is to serve as a contact between the psychologists and the public. When a business man has a problem of a psychological nature the Corporation stands ready to consider his problem and refer it to some reputable psychologist who is qualified to deal with it. The value of this procedure consists in preventing the business man from purchasing a gold brick or employing some self-styled psychologist who is inadequately trained and will probably do more harm than good. Such inquiries are carefully considered and turned over to some one competent to handle them, frequently through some of the branches in different States. The Corporation is also devising various standardized tests which may be given in its different branches as a routine procedure. Some of these will be of the clinical type and some more definitely aimed at vocational guidance. All profits of the Corporation after expenses have been met and the overhead paid are, according to the charter, to go to research purposes. This suggests a similar policy of some business concerns. The telephone companies, for instance, charge a slight margin above the actual cost and this additional fund is devoted to improving telephone service by research work. An electric lamp company will frequently charge somewhat more than the price of the commodity in order to have money to experiment further and devise better lamps. So with the Psychological Corporation profits on consulting work are devoted to further studies which will produce greater welfare in the end. Perhaps the most important work of the Corporation at present, however, is that above mentioned of connecting a person needing psychological advice with some psychologist competent to give it. A business man naturally needs some one to advise him in the selection of research personnel because he is not familiar with psychology and has no means of evaluating those who call themselves psychologists. The Psychological Corporation aims to provide such advice.

### ATTITUDE OF WORKERS AND MANAGEMENT TOWARD EMPLOYMENT PSYCHOLOGY

The foregoing gives some notion of the outstanding trends of employment psychology at the present time. The various indi-

vidual scientists are doing their part and larger organizations are making a definite contribution toward advancing the status of applied psychology in general and employment psychology in particular. The success of this program depends, however, to a considerable extent, on the attitude of those involved.

**Workers.** The attitude of the workers toward employment psychology has not manifested itself very definitely or unmistakably as pro or con. Of course there is a natural suspicion of any innovation that apparently aims at efficiency. There have been many instances where methods of scientific management have been misused, not through any fault of the principle, but because of the abuse of the practice. Employees have observed improvements brought about by such methods without any measurable benefit to themselves and they have naturally been disgruntled. This attitude has not spread seriously to mental tests as yet. Some applicants taking such tests seem quite interested, others take it as a matter of course, and a relatively small number become distinctly disgruntled and express their opinion in a forceful way that they consider this an undesirable method of getting a job.

An impartial consideration of the foregoing chapters will indicate that such a hostile attitude is ungrounded. Employment psychology aims to benefit the employee as well as the employer. It may seem hard in a given instance to reject a particular man who needs a job, but it is probably a kindness to him in the long run not to hire him for a job in which he has no possible future, thereby decreasing his chances of getting into work for which he is adapted. The man who is placed in a job for which he has the aptitude will enjoy his work and will in general be happier. Applied psychology is distinctly impersonal. It aims to discover the facts and derive methods regardless of who uses them. The psychologist could just as well be retained as a consultant by the workers as by the management. Theoretically a factory operated by a council of employees should be just as enthusiastic about psychology as a factory operated on the present basis. It is desirable to educate the workers to realize the impersonal character of employment psychology. They should be made to see the desirability of not giving a man a wage and a job arbitrarily, but of discovering and

developing his particular ability to best advantage. There is no waste so far-reaching as misdirected human activity, and waste in industry hits all of us, including the worker himself.

It is encouraging to note in Europe an increasing realization on the part of labor that the union itself suffers from vocational misfits. In Berlin the trade unions actually contribute to the support of a bureau which functions especially in selecting apprentices. Resolutions have been passed by some of the trade unions in favor of psychological vocational guidance. The Krupp workers have a psychologist to study the industrial applications of the science. In this country the American Federation of Labor is represented in the Personnel Research Federation.

**Management.** The attitude of the management toward employment psychology is likewise important. While some executives still feel self-sufficient in dealing with the human element, the majority are coming to realize their own limitations or are at least willing to submit their own opinions to scientific evaluation. They must, moreover, appreciate the scientific attitude and the necessity for investigating minutiae, for repeating observations again and again and for amassing statistical data. They must consider the general results rather than the individual case which may be an exception to the rule. When dealing with vocational prediction it is a question of probabilities, and even though the methods are rather successful there are bound to be some erroneous predictions. The executives must learn to consider the proportion of successful placements rather than the results with a single man. Finally, they must be patient with the slow, painstaking character of scientific research.

#### NECESSITY FOR FURTHER RESEARCH

Granted that workers and management are willing to cooperate in developing psychological methods for employment, it is scarcely necessary to stress the importance of further research in this field. We obviously need more facts, and we cannot determine whether a given procedure is the proper one until we try it out. Industries realize the importance of research in other technical lines. They hesitate to base decisions upon opinion when facts may be obtained.



Many concerns, of course, maintain their own physical or chemical laboratories. Research in psychology is often just as important as in these other sciences. While a concern will measure the specific gravity of certain compounds used in its products, it is less inclined to measure the mental capacities of the workers who are handling those products. It is impossible to solve these problems by intuition just as it is impossible to determine the weight of a liquid by looking at it.

**Problems of individual concerns.** Much of the research that is necessary grows out of the individual problems of a given plant. Each concern will frequently have its own special situations which need specific study. We cannot take a technique developed in one field bodily into another without evaluating it in the latter situation. The trade tests, for instance, which proved useful in the army have not in every case proved successful in industries because the work of a particular tradesman in the army was somewhat different from that of a similar tradesman in a particular industrial concern. Clerical tests developed in one case may be unsatisfactory for use in another because computing machines are used in the first instance but not in the second. A rating scale developed in one organization would not necessarily work out well in another, for the first concern might be rating one kind of executive and the second concern a distinctly different kind. Each individual concern must then validate the psychological methods in its own situation before putting them into practice. Even with tests that have been rather well standardized and put out in commercial form, it is well to make a preliminary study of them in the new situation before attaching too much value to them. Any individual concern thus presents a variety of problems for psychological research.

**Special occupations.** In addition to research of the above type in validating a previously developed method in a given plant or in devising new methods for the local conditions, there are other problems of a more general nature with which employment psychologists must concern themselves — problems to which many individual workers must doubtless contribute before they are finally solved. For instance, if one goes through the gamut of occupations he will find some in which satisfactory experimental results have been ob-

tained and others in which apparently little has been accomplished. Clerical workers, for example, have been rather extensively studied. This type of work apparently necessitates certain rather specific capacities which are objectively measurable. The ordinary clerical worker requires a certain alertness, skill with the fingers, ability to deal with numbers, to classify topics, to detect errors in spelling, and the like. These capacities and abilities have been measured by various tests. A somewhat similar situation has been found with various factory operations. A given job requires perhaps a particular kind of coördination between eye and hand, a certain reaction time or type of attention — processes which can be measured by conventional test procedure. Furthermore, in the case of clerical or industrial workers it has frequently been possible to find a considerable number of persons doing the same sort of work on whom the test can be standardized.

The situation is quite different when we deal with such complex things as executive ability. The executive has to reason, make decisions, deal with men, cooperate, get things done, delegate authority, and the like. These traits or capacities are not so readily measurable as are those required by clerical or industrial workers. We have up to the present approached them largely through the technique of rating scales, and progress has not been rapid. More objective methods will be necessary before the problem of selecting executives is satisfactorily solved. Moreover, if the measurements themselves are perfected, it will often be difficult to validate them for the reason that it is unusual to find a considerable number of executives doing approximately the same thing. In a factory we may very readily find a hundred men building the same kind of automobile tire, but, if we should select a hundred executives from the same concern, we should probably find them doing approximately one hundred different things, so that it would be more difficult to obtain the criterion by which to evaluate our measurements. However, as time progresses it will doubtless be possible to select certain aspects of executive ability which are rather common to a great many positions and devise methods for measuring those particular aspects.

There are other occupations that are in somewhat this same

status. Salesmanship, for instance, has been studied to quite an extent, but the problem of selecting salesmen has by no means reached its final solution. The characteristics which constitute a successful salesman are apparently exceedingly complex, many of them involving personality traits rather than mental capacities. While various ingenious tests have shown some indication of selling success and while various items of personal history have been somewhat differential, there still remains a great deal of research to be done in this field. In the various professions there has been very little research indeed. The business man is, of course, not so much interested in selective methods for the professional fields except possibly in engineering. However, the development of vocational standards for all lines of work is a step in the whole program of adjusting people more satisfactorily to the type of work for which they are best fitted.

**Special techniques.** Another field for general research contribution lies in the development of further mental measurement techniques. We already have fairly satisfactory tests for some of the simpler capacities and abilities. While these are undoubtedly of great importance in many occupational lines, nevertheless any psychologist realizes that other things are also necessary. It is not always a question of what a workman can do, but of what he will do. His attitude toward his work and the way he approaches it are important considerations in his occupational prognosis.

We have seen earlier that certain preliminary methods have been devised to measure interest as well as ability. The work in this field has only begun, but ultimately we shall have fairly well standardized methods for determining a person's vocational and avocational interests with a view to placing him in some position where these interests will facilitate rather than hinder his progress.

Then there is the whole field of temperament and personality which has barely been touched upon as far as actual measurement is concerned. We need better methods of evaluating such things as honesty, flexibility, stick-to-it-iveness, adaptability, tact, enthusiasm, and the like. At present we have systematized efforts to rate such qualities, but this usually necessitates acquaintance of the rater with the person who is to be rated. What we need and what

further research may give us are objective methods of measuring these things in the same fashion that we can measure intelligence or memory or reaction time.

Even in the field of intelligence measurements it will be recalled that three types of intelligence have been suggested, verbal or abstract, mechanical, and social. Most of the work hitherto has dealt with the first of these. We are at present in the midst of considerable research upon the second, but the third has scarcely been touched at all. A field for much needed research lies in the development of measurements of this social intelligence in order to determine a person's general ability in dealing with a social situation as compared with his ability in dealing with more abstract things. This technique will be especially valuable in employment problems dealing with occupations where the individual makes very definite social contacts and where his success in the occupation depends somewhat upon his adaptability in making such contacts.

These, then, are some of the outstanding problems for employment research in the immediate future. In addition to specific local problems in individual plants there is much to be done by various research workers in further studying those occupations which require more complex and less tangible mental characteristics and in perfecting techniques for objectively measuring character, personality traits, and social intelligence.

#### **ESSENTIALS FOR FUTURE RESEARCH**

**Competent psychologist to conduct the research.** The foregoing are some of the problems with which research workers in employment psychology must in the future concern themselves. We will now consider some of the conditions necessary for successful research work in this field. In the first place, a competent psychologist should be obtained to conduct a given piece of research. Earlier chapters have indicated that this type of work involves rather special technique and requires a person with some experience in mental measurements and some appreciation of individual differences. After measurements have been put into final form so that they are relatively fool-proof, it is then time to turn them over to untrained individuals for routine administration. Even then there is some-

thing to be said for the value of a modicum of psychological training for those who administer tests and interpret the results. But in the process of developing methods before their final application, laboratory training is invaluable. In such a research program contingencies are apt to arise which would lead the untrained experimenter into various errors. He might fail to establish *rappor*t, fail to control the attention of the subjects, be uncertain what to do in the case of a bad start, and overlook various incidental reactions of the subjects which might still be of considerable significance. A concern would not put into its industrial laboratory a chemist who had had no laboratory experience, but had merely taken theoretical courses and read about the subject. He would be liable to drop the test-tubes, mix the stoppers of the reagent bottles, and punch a hole in the filter paper. Similarly a psychologist without laboratory experience would be inclined to vary the test instructions, to overlook various conditions of illumination, and the like, to fail to eliminate unnecessary distractions during the test and to be careless with the temporal aspects of the procedure. Aside from the mere conduct of the test the laboratory background gives one a scientific attitude in interpreting the results. The uninitiated is apt to stress some aspects that appeal to him. His grading of a test blank may frequently be colored by his general impression of what the man ought to do so that he will grade him too leniently or too stringently.

The importance of obtaining a competent psychologist is stressed because there have been instances in which business men employed persons who purported to be psychologists, but who were not adequately trained. These individuals were naturally unsuccessful in their practical work and to some extent this brought discredit upon the science in general. While many individuals like this were not fraudulent in any sense of the word, they were nevertheless incompetent and should not have been engaged in that type of work. As suggested earlier in the chapter one means of ascertaining whether an individual is competent for such work is through the Psychological Corporation which endeavors to connect persons needing psychological service with some one who can adequately perform that service.

## THE OUTLOOK FOR EMPLOYMENT PSYCHOLOGY 505

**Adequate criteria.** A second essential for future employment research is adequate criteria. In Chapter VI the fact was stressed that final psychological measurements can be no more valuable than the criteria by which they are evaluated. Obtaining such criteria depends on the coöperation of all those concerned in furnishing such data. If foremen or managers or others are called upon to rate the men under them in some fashion, it is essential that they take this work seriously and make the ratings with the greatest possible care. With reference to production criteria, of course the research depends upon full access to all production records that are available. If such production records are to be valuable, they must obviously have been accurately kept.

**Subjects on whom to standardize methods.** A third essential for such research is the subjects on whom the experiments are to be conducted. Access must be had to employees (or possibly applicants) on whom to standardize the various measurements. The psychologist must go right into the plant with his tests and measurements. He could not standardize, for instance, a vocational test for lathe operators on students in an Arts College. He must actually evaluate it with men who are in the practical work. This may cause some inconvenience at the plant where the research is being done, but it is nevertheless necessary. Furthermore, the subjects who are used must coöperate and do their best in taking the tests. The only way to keep incentive constant, as has already been suggested, is to keep it at a maximum. Results will naturally be meaningless if one employee does his best and another does not try. Consequently, such a program cannot be carried through successfully where the morale is low and the persons taking the tests are unwilling to coöperate. In addition to obtaining employees who are willing to do their utmost, it is further necessary to have enough of them to make the results reliable. The psychologist cannot be expected to solve the problems of a given vocation by having six men sent to him for examination. According to the general principle of averages the more that are included the more apt are the results to represent typical tendencies.

**Facilities for conducting research.** A fourth essential for personnel research is adequate facilities for conducting the work. In

giving a test, for instance, it is essential that all the subjects have approximately standard conditions. It would be impractical to test some persons in the shop and some in the laboratory because of the different amount of distraction. A separate laboratory is presumably desirable where lighting, ventilation, and other external conditions can be kept in an optimum condition. Adequate time should be allowed, moreover, for each subject who is tested. If it becomes necessary to rush, the examiner is liable to make various errors himself and his attitude of excitement is quite apt to be communicated to the subjects. Consequently, one would not be enthusiastic about testing a group of men at lunch hour or after the day's work. It is usually necessary to test the men some time during working hours on the company's time in order to insure standard conditions and adequate time for each as well as to insure proper morale.

**Opportunity to evaluate results adequately.** A fifth requisite is opportunity to study the results adequately without pressure. An executive is liable to consider the psychologist as he does his salesman and look for immediate returns. It is unwise to crowd a research worker. Discoveries cannot be made to order. The management must be patient with the research department. Sometimes they naturally go into blind alleys and must start over again. But if one considers the number of reagents that were tried before the discovery of the one which when mixed in gasoline would eliminate the knock, he will be inclined to pardon an employment psychologist for making a few false starts that do not lead directly to the mark. Scientific facts do not spring up overnight. A thing that irritates a research worker more perhaps than anything else is pressure to uncover fundamental truths on schedule. In this connection the research worker should have ample opportunity to follow up his results. He may have devised a set of measurements which apparently indicate aptitude for some particular line of work. He may not be fully satisfied, however, with the results until he has checked them on a new group of people who are selected on the basis of such measurements and who subsequently demonstrate their fitness or unfitness. Such subsequent validation of the measurements should by all means be permitted and encouraged.

## THE OUTLOOK FOR EMPLOYMENT PSYCHOLOGY 507

**General coöperation.** The research worker finally requires the general coöperation of all those with whom he comes in contact. The scientist will not do his best if there is some one continually opposing him. His own morale should be considered as well as that of the workers. He often needs advice on many points, he requires records and supplies, clerical assistance is sometimes necessary, and various accommodations may be made for him in the shifting of schedules or in providing as subjects some particular group of workers who are of especial interest. Every one with whom he is working must be very definitely "with him" in the project. It is probably preferable for him to be considered, temporarily at least, as an integral part of the staff or at least to have his status in the organization very definitely recognized.

### THE SOCIAL IMPLICATIONS OF EMPLOYMENT PSYCHOLOGY

Before concluding the discussion of the outlook for employment psychology, we should consider again its broad social implications brought out at the end of the first chapter. The methods described in the preceding pages can be of as much benefit to the employee as to the employer. It is really a kindness to an applicant not to hire him for a job in which he has little chance of success, because he thereby has a greater probability of locating something in which he has a future. Misdirected human activity is one of the greatest wastes in our civilization and it indirectly affects all of us. Furthermore, while the techniques discussed above are for the most part objective, impersonal, and statistical, this does not mean that the employment process should be stereotyped and mechanical. As suggested in Chapter I, we must after all regard the applicant as an individual who has certain capacities, but likewise certain interests and who is looking for opportunities. His interests must be treated with respect and tact especially if they are apparently at variance with his capacities. The technique of employment should be tempered with a certain amount of common sense and appreciation of the unique problems of the individual. He should be aided so far as possible in finding himself and in improving his opportunity. But after taking these things into consideration the major part of the problem still consists in measuring the potentialities which the man



brings from his ancestry to the employment office and comparing them with objective standards that have been developed for the particular jobs in question. This is the largest contribution which psychology has to make to the increasing of human efficiency by scientific selection of personnel.

Efficiency, however, should not be achieved at the expense of happiness nor should happiness be obtained at the expense of efficiency. The happiness to be considered, however, is ultimate rather than immediate — the real happiness that comes from the expression of normal cravings for achievement, freedom from fear or jealousy, reasonable leisure, and a sense of accomplishing something worth while. From this standpoint we should consider the capacities and the interests of the man and attempt to adapt him to his work and adapt the work to him so that that unit will be of maximum effectiveness. Employers are often apt to shy at the notion of happiness as one of the goals for scientific effort. Some of them doubtless have had unfortunate experiences with professional uplifters. The psychologist, however, is not thinking in these terms, and he is not short-sighted in his belief that a happier society is a more effective society. It is difficult to say how much of our industrial unrest and unhappiness is due to the maladjustment of the worker to his work. The cause of the unrest, actually voiced, is often not the real cause. In many instances persons have been known to protest about their wages when the real thing that was bothering them was the climate. They may apparently find disagreeable aspects in their working conditions when the real fault is that they are individually not adapted to their work.

In this scheme of things applied psychology will in the future play an increasingly large rôle. The science has, to be sure, been "oversold" in a few instances. It is a rather common tendency among business people and others to claim too much for something which they have to sell, and psychology immediately after the War with its surplus enthusiasm went perhaps a little too far in this respect. But the lean years of the business cycle purified its soul. We have gone back again to fundamentals and are proceeding with painstaking and thorough scientific procedure.

Personnel research is a comparatively new study, mental measure-

## **THE OUTLOOK FOR EMPLOYMENT PSYCHOLOGY 509**

ment is not well known and appreciated by the layman, and it will take considerable time before people come to appreciate these things fully. It took a long time, for instance, to remodel our social attitude toward crime. The same thing will doubtless be true of the social attitude toward applied psychology in general and employment psychology in particular.

The broad movement to study man has just begun. Psychology is now playing an increasing rôle in the school, in the clinic, in the advertising agency, in the factory, and in the employment office. These problems of life adjustment are coming more and more to the front. The last century was characterized by tremendous advances in the natural sciences and in the technologies. The present one bids fair to be an era for human engineering. The psychologist's ideal is to have every one provided with the opportunity to do that particular part of the world's work for which he is best adapted and in which he is most interested. When this ideal is achieved, the world will be a happier place for all of us.

## APPENDIX I

### ILLUSTRATING THE TECHNIQUE OF CORRELATION

THE notion of correlation is fundamental in employment psychology. We are often concerned with the extent to which two variables or sets of traits or measurements are related. We may wish to determine whether or not estimates of a trait made by acquaintances are at all related to estimates made by unacquainted persons judging purely from physiognomy. We may desire to ascertain to what extent efficiency in a particular mental test is related to efficiency in a job. The ultimate aim is usually to predict one variable in terms of another; hence the need for expressing quantitatively the relation between the two variables. The correlation coefficient is the standard technique for expressing this relation. The present section aims merely to give a simple notion of how correlations are obtained and the meaning of correlations of different magnitudes. The examples cited are made absurdly brief in the interest of avoiding tedious arithmetical computations. With longer examples, the arithmetical procedure would naturally be more arduous. There are available, however, various short-cut procedures which are described in books on statistics. (498, 603.)

One of the simplest correlation procedures is that involving rank differences. Given two series of measures it is possible to rank them both and get the differences in the rank. Consider Example I, which gives data on five men. Let us suppose that these five men make the scores in a mental test indicated in the first column and that some quantitative statement of

EXAMPLE I

NAME	TEST SCORE	JOB SCORE	TEST RANK	JOB RANK	RANK DIFFERENCE	RANK DIFFERENCE SQUARED
Adams . . . . .	45	75	1	3	2	4
Andrews . . . . .	43	80	2	2	0	0
Briggs . . . . .	38	82	3	1	2	4
Brown . . . . .	36	63	4	5	1	1
Doe . . . . .	34	68	5	4	1	1

Sum of rank differences squared 10

$$r = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} \qquad \begin{matrix} \sum D^2 = 10 \\ N = 5 \end{matrix}$$

$$r = 1 - \frac{6 \times 10}{5(25 - 1)} = 1 - \frac{60}{5 \times 24} = 1 - .50 = .50$$

their ability in the job, such as units of production, gives the scores indicated in the second column. The problem is the extent to which those who make high test scores make high job scores and *vice versa*. It is to be noted that Adams makes the highest test score and is given the rank of 1 (cf. column headed "test rank"); Andrews is next best in the test and is given the rank of 2; Briggs comes third. If, now, we consider their job scores, Briggs is the best of the group and is ranked 1 (cf. the column headed "job rank"); Andrews is second best and gets a rank of 2, while Adams falls in third place. We may now neglect the first two columns and consider only the columns of ranks and determine the difference of rank in each instance. Adams is ranked 1 on the test and 3 on the job and the difference between these figures is 2 (cf. the column headed "rank difference"). Andrews is ranked two in both cases, so the difference is 0; Briggs is ranked 3 on the test and 1 on the job and the difference is 2. Similarly, the differences for Brown and Doe may be computed. These differences give some notion of the extent to which the two series of ranks correspond. If the difference is as great as 4 or 5, it indicates that a person is ranked in one trait very differently from the way in which he is ranked in the other. If the difference is small, it indicates that there is a fair correspondence. In working out the correlation coefficients on the basis of these data, it is necessary to square the rank differences as is done in the last column. The sum of these squares is then obtained. The formula for computing the coefficients is indicated in the example. It must be taken on faith, in the present connection, but its derivation may be obtained in various works on statistics. (692.) In the formula the term  $\sum D^2$  means the sum of the differences squared, while  $N$  means the number of cases involved — in this instance, 5 men. To solve the formula it is necessary to take 6 times the sum of the differences squared, divide this by  $N$  times  $N^2 - 1$  and then subtract this quotient from 1. In the present example this works out to an answer of .50. It is rather conventional procedure to carry correlation coefficients out to two decimal places. This particular coefficient indicates a fair degree of correlation, which, of course, is obvious from inspection of the original data, but it would not be so obvious if a large number of individuals had been involved. For purposes of comparison several other examples are added using similar data that give higher or lower correlations than those of Example I. Example II, for instance, indicates a case of what is termed perfect correlation. Briggs, who ranks highest in the test, is also highest in the job. Andrews, who is second in the test, is second in the job, and so on down to Doe, who is poorest in each respect. In this case there are no differences in rank and the correlation coefficient comes out 1.00, which is the maximum possible. This indicates a perfect correspondence between the two variables.

Example III presents a perfect negative correlation. Adams, who is best in the test, is worst in the job, and Andrews, who is second best in the

test, is second worst in the job, and so on down to Doe, who is worst in the test and best in the job. This, of course, makes the differences in rank as large as possible and the coefficient is  $-1.00$ , which is the maximum possible negative coefficient. It indicates a perfect tendency for the highest scores in one variable to go with the lowest scores in the other.

## EXAMPLE II

NAME	TEST SCORE	JOB SCORE	TEST RANK	JOB RANK	RANK DIFFERENCE	RANK DIFFERENCE SQUARED
Adams.....	38	74	3	3	0	0
Andrews.....	43	79	2	2	0	0
Briggs.....	45	82	1	1	0	0
Brown.....	36	68	4	4	0	0
Doe.....	33	62	5	5	0	0

$$\Sigma D^2 = 0$$

$$N = 5$$

$$\rho = 1 - \frac{6 \times 0}{5(25 - 1)} = 1 - \frac{0}{5 \times 24} = 1.00$$

## EXAMPLE III

NAME	TEST SCORE	JOB SCORE	TEST RANK	JOB RANK	RANK DIFFERENCE	RANK DIFFERENCE SQUARED
Adams.....	46	67	1	5	4	16
Andrews.....	43	64	2	4	2	4
Briggs.....	38	75	3	3	0	0
Brown.....	36	78	4	2	2	4
Doe.....	34	83	5	1	4	16
						<u>40</u>

$$\Sigma D^2 = 40$$

$$N = 5$$

$$\rho = 1 - \frac{6 \times 40}{5(25 - 1)} = 1 - \frac{240}{5 \times 24} = 1 - 2 = -1.00$$

Example IV involves a correlation of  $.80$ , which is not perfect, although very high. There are slight discrepancies in rank, enough to spoil the perfection, but from inspection it is obvious that there is a very close relation between the variables and this is reflected in the high coefficient. Example V indicates a 0 correlation — that is, a situation in which there is no apparent relation between the two variables. Inspection reveals that it

would be practically impossible to predict a man's job rank if his test rank were known. This is reflected in the coefficient of 0.

## EXAMPLE IV

NAME	TEST SCORE	JOB SCORE	TEST RANK	JOB RANK	RANK DIFFERENCE	RANK DIFFERENCE SQUARED
Adams.....	40	75	3	3	0	0
Andrews.....	48	79	1	2	1	1
Briggs.....	44	85	2	1	1	1
Brown.....	32	68	5	4	1	1
Doe.....	37	64	4	5	1	1
						4

$$\Sigma D^2 = 4$$

$$N = 5$$

$$\rho = 1 - \frac{6 \times 4}{5(25 - 1)} = 1 - \frac{24}{5 \times 24} = 1 - .2 = .80$$

## EXAMPLE V

NAME	TEST SCORE	JOB SCORE	TEST RANK	JOB RANK	RANK DIFFERENCE	RANK DIFFERENCE SQUARED
Adams.....	43	86	2	1	1	1
Andrews.....	48	73	1	4	3	9
Briggs.....	37	68	4	5	1	1
Brown.....	39	77	3	3	0	0
Doe.....	31	80	5	2	3	9
						20

$$\Sigma D^2 = 20$$

$$N = 5$$

$$\rho = 1 - \frac{6 \times 20}{5(25 - 1)} = 1 - \frac{120}{5 \times 24} = 1 - 1 = .00$$

The method of rank differences, while a very convenient and relatively easily computed correlation procedure, is not ideal because it assumes that the differences between any two adjacent ranks in one variable are all equal. Referring to Adams's test score in Example I, it is actually two points better than Andrews's, while the latter is five points superior to Briggs's, but in ranking them it is assumed that these differences are equal. In this way a striking superiority or inferiority of some individual may be overlooked. A standard method is available which takes into account the

actual magnitude of the scores. Instead of merely considering whether a person who is best in the test is best in the criterion, we are concerned with whether a man who deviates from the average in one respect deviates correspondingly in the other. In Example VI an illustration is given of the

EXAMPLE VI

NAME	TEST SCORE	CRITERION	DEVIATIONS		DEVIATIONS SQUARED		PRODUCT OF DEVIATIONS
			Test	Criteria	Test	Criteria	
			X	Y	x	y	
Adams.....	10	15	+4	+6	16	36	+24
Andrews.....	7	8	+1	-1	1	1	-1
Briggs.....	5	6	-1	-3	1	9	+3
Brown.....	2	6	-4	-3	16	9	+12
Doe.....	6	10	0	+1	0	1	0
Total.....	30	45			34	56	+39-1
Average.....	6	9			6.8	11.2	+38
					2.61	3.34	

$$r = \frac{\sum xy}{N \sigma_x \sigma_y} \qquad \sum xy = 38 \qquad \sigma_x = 2.61$$

$$\qquad \qquad \qquad N = 5 \qquad \qquad \qquad \sigma_y = 3.34$$

$$r = \frac{38}{5 \times 2.61 \times 3.34} = \frac{38}{43.6} = .87$$

computation of correlation by the so-called "products-moments" (that is, products of deviations) method. The first part of the computation is identical with that previously described in connection with standard deviation (p. 161). The original scores are given in the first two columns. Conventional procedure calls these original scores X and Y. The average of each column is computed. The third column gives the deviations of each test score from the average test score. Adams's score of 10 is 4 greater than the average of 6, that is, its deviation is +4; Andrews's test score of 7 is 1 above the average, while Briggs's is 1 below, etc. Deviations of criterion scores from the average are computed similarly. The deviations are denoted by x and y. The deviations are now squared and each column averaged. The square roots of these averages give standard deviations of 2.61 and 3.34 respectively.  $\sigma_x$  denotes the standard deviation of the test scores and  $\sigma_y$  the standard deviation of the criterion. The next step is to take the product of the deviations. For instance, Adams's deviation of +4 in the test is to be multiplied by the corresponding deviation of +6 in the criterion, giving a product of 24. Andrews's figures are +1 and -1 respectively and the product is -1. These products are then

algebraically totaled, giving 38, and we are ready to substitute in the formula.  $\sum xy$  denotes the sum of the products of the deviations of the variables — in this case 38;  $N$  denotes the number of individuals, and  $\sigma_x$  and  $\sigma_y$  the standard deviations of the two variables as above described. Substituting in the formula the coefficient of .87 is obtained. This has taken into account the actual magnitude of the original measures and not merely their relative standing. It is to be noted that if a given individual's measures are both above or both below the average, the product of the deviations will be plus and the numerator of the fraction in the formula large, while if one is above the average and the other below, the product will be negative and the sum of the products will be somewhat decreased and the coefficient lowered. This type of coefficient gives probably the best indication of the relation between the two measures and is very widely used in all the most careful statistical work.

The foregoing example may be misleading as to the simplicity of the arithmetical work involved. When there are 50 or 100 individuals in the correlation, the work of computation becomes very considerable. In such cases there are, however, various short-cut procedures which decrease very appreciably the routine work. It is beyond the scope of the present discussion to describe these. Some involve grouping the data into classes rather than using the original scores, some are designed to obviate much of the multiplication involved in getting the  $xy$  term; some obviate the necessity of computing deviations at all, and others are especially devised for use with some particular type of computing machine. Some of these methods are discussed in current statistical works. (498, 603.)



## APPENDIX II

### COMBINING RANKINGS WITH INCOMPLETE DATA

THE need was mentioned in Chapter VI of a technique for combining incomplete rankings. If, for instance, several foremen independently rank a group of workers as to their ability in the job, but some foremen rank all the workers and others rank only a limited number, it is inadvisable to average the ranks assigned each worker in order to determine his general standing. A rank near the bottom of the scale made by a foreman who estimates all the workers contributes a more severe penalty to the average than does a rank near the bottom of the scale by a foreman who estimates a smaller number. (Cf. p. 159.) This same problem may arise in rating scale procedure, in evaluating application blanks, recommendations, letters of application, or the results of employment interviews, or in any other situation where a number of judges make an estimate of something and pool their results.

Various methods for combining these incomplete rankings have been suggested. (Cf. 599, 299.) The one to be described is probably as simple and satisfactory as any. (473.) It assumes that the ranks assigned by any foreman or other judge conform to a normal frequency curve. When dealing with quantitative ratings rather than ranks (cf. Chapter VI), a similar assumption was made, and each estimate by a given foreman was converted into its deviation from the average estimate made by that foreman divided by the standard deviation of his estimates. Such converted estimates made by different foremen were then comparable because they were all located on normal frequency curves and the properties of such curves are universal.

These same principles are applicable when the estimates are in the form of ranks. Suppose five workers have been ranked by a foreman, and it is desired to convert these ranks into terms of standard deviation ( $\sigma$ ). By consulting tables (derived originally by calculus) we find that in a normal frequency curve the average of the highest 20 per cent of the individuals is theoretically 1.40  $\sigma$  above the average score or estimate; the average of the next 20 per cent is .53  $\sigma$  above the average, while the average of the middle 20 per cent coincides with the average of all, so its deviation is 0  $\sigma$ . Similarly, the average of the next 20 per cent of the individuals is .53  $\sigma$  below the average or  $-.53 \sigma$ , while the average of the lowest 20 per cent is  $-1.40 \sigma$ . We may then replace the original ranks by these converted figures, thus expressing them in terms of standard deviation and making all such ranks comparable regardless of the number of workers that is ranked. A table

may be derived based on the above considerations which gives similar figures for converting rankings that involve different numbers. A portion of such a table is reproduced in Table LXXIII. It could be carried out to include any desired number of individuals.

TABLE LXXIII. FOR CONVERTING RANKS INTO TERMS OF STANDARD DEVIATIONS <sup>1</sup>

RANK	NUMBER OF INDIVIDUALS RANKED								
	2	3	4	5	6	7	8	9	10
1	.80	1.10	1.27	1.40	1.49	1.59	1.67	1.71	1.76
2	-.80	0	.32	.53	.68	.81	.91	.98	1.04
3		-1.10	-.32	0	.22	.37	.59	.60	.68
4			-1.27	-.53	-.22	0	.18	.29	.39
5				-1.40	-.68	-.37	-.18	0	.13
6					-1.49	-.81	-.50	-.29	-.13
7						-1.59	-.91	-.60	-.39
8							-1.67	-.98	-.68
9								-1.71	-1.04
10									-1.76

<sup>1</sup> After Ream.

In using the table, if a given foreman has ranked five men, we look in the column headed 5. To the man whom the foreman ranked first we give a value of 1.40, to the man whom the foreman ranked second a value of .53, and so on. Suppose the next foreman has ranked only four men. We then use the column headed 4 and give his first man a value of 1.27. To illustrate the procedure a simple example is given in Table LXXIV. The left portion of the table gives the original ranks assigned to the workers by three foremen. The third foreman, however, ranked only four of the workers. The next portion of the table gives these same ranks converted into terms of standard deviation by consulting the previous table. The last column gives for each man the average of his converted ranks. It may be noted that if the original ranks had been averaged without conversion Brown and Doe would have the same average of four. With the converted ranks, however, Brown is somewhat inferior to Doe.

## APPENDIX

519

TABLE LXXIV. ILLUSTRATING COMBINATION OF INCOMPLETE RANKS BY CONVERSION INTO TERMS OF STANDARD DEVIATION

	ORIGINAL RANKS			CONVERTED RANKS			
	First foreman	Second foreman	Third foreman	First foreman	Second foreman	Third foreman	Average of converted ranks
Adams . . . . .	1	2	1	1.40	.53	1.27	1.07
Andrews . . . . .	3	1	2	0	1.40	.32	.57
Briggs . . . . .	2	4	4	.53	-.53	-1.27	-.42
Brown . . . . .	4	5	3	-.53	-1.40	-.32	-.75
Doe . . . . .	5	3		-1.40	0		-.70

Other schemes for handling this sort of data are available. One of them provides a table similar to the preceding, but the results are in terms of something analogous to percentile scores. This obviates the use of decimal points and minus signs. (225.) The method above described, however, is typical of those used for combining incomplete ranks.

### APPENDIX III

#### ILLUSTRATING THE DERIVATION OF A REGRESSION EQUATION

In Chapter IX we saw the importance of partial correlation and the regression equation for weighting a number of tests in order to get the best possible prediction of vocational aptitude. A brief indication of the technique was given in that connection. In the present section a regression equation in four variables is worked out in detail by way of illustration of the process. The equation is that already mentioned on p. 242 for predicting ability at finishing tires on the basis of three tests — cancelling adjacent pairs of numbers whose sum is 10, finding consecutive numbers arranged irregularly in a square, and simple visual reaction time. These tests in the previous connection were numbered 3, 5, and 7. In the present section, in the interest of simplicity, they will be renumbered respectively 2, 3, and 4.

The original correlations of the tests with the criterion and with each other (cf. Table XXII) are as follows, where 1 is the criterion and 2, 3, and 4 are the tests just mentioned. For instance,  $r_{12}$  — that is, the correlation

	2	3	4
1	.51	.49	-.41
2		.66	-.24
3			-.22

between the criterion and test 2 — is .51, while  $r_{13}$  is .49 and  $r_{14}$  is -.22. From these correlations, which are termed “zero order” coefficients, it is possible to derive any coefficient of the “first order” like  $r_{12.3}$ , which means the correlation between the criterion and test 2 with test 3 kept constant. Such coefficients are called “first order” because there is one secondary subscript, that is, one subscript after the point or one variable that is kept constant. There are always two primary subscripts before the point. From coefficients of the first order it is possible to derive those of the second order like  $r_{12.34}$ , which have two secondary subscripts or two variables that are kept constant. From these we may derive third order coefficients such as  $r_{12.345}$ , etc.

The formula used in all such computations is of the form:

$$r_{12-3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}$$

The subscripts of the first term in the numerator, it is to be noted, are the same as the primary subscripts of the coefficient for which we are solving. (12.) The second term in the numerator is the product of two factors. Each of these has a subscript (3) the same as the secondary subscript of the coefficient for which we are solving. The primary subscripts of this coefficient for which we are solving appear also in this second term, one in each factor. Putting it in another way, we obtain the subscripts of this second term by combining the secondary subscript of the coefficient for which we are solving (3) first with one of its primaries (1) and then with the other (2). The two subscripts that appear in the denominator of the formula are identical with those in the second term of the numerator.

If we substitute in this formula the zero order coefficients given in the table above we have:

$$\begin{aligned} r_{12-3} &= \frac{.51 - .49 \times .66}{\sqrt{1 - .49^2} \sqrt{1 - .66^2}} = \frac{.51 - .324}{\sqrt{.759} \sqrt{.564}} \\ &= \frac{.186}{.87 \times .75} = \frac{.186}{.653} = .28 \end{aligned}$$

This tells us that the correlation between the criterion and test 2 would be .28 if we had persons with identical ability in test 3. In exactly the same way the other coefficients of the first order may be derived. For instance:

$$r_{24-3} = \frac{r_{24} - r_{23} r_{34}}{\sqrt{1 - r_{23}^2} \sqrt{1 - r_{34}^2}}$$

Here again it is to be noted that the subscripts of the first term in the numerator are the primary subscripts of the coefficient for which we are solving (24), these same primaries appear in the second term — one in each factor — while the secondary subscript (3) appears in both factors and the subscripts in the denominator are the same as those in the second term in the numerator. Substituting the zero order coefficients in this formula:

$$\begin{aligned} r_{24-3} &= \frac{-.24 - .66 \times (-.22)}{\sqrt{1 - .66^2} \sqrt{1 - (-.22)^2}} = \frac{-.24 - (-.145)}{\sqrt{.564} \sqrt{.952}} \\ &= \frac{-.095}{.751 \times .975} = \frac{-.095}{.732} = -.13 \end{aligned}$$

In this same manner all the coefficients of the first order may be computed. In the present problem they are not all necessary, but those that are required for subsequent use are as follows. Their method of derivation is identical with the preceding.

$r_{12-2}$	.28	$r_{14-2}$	-.34
$r_{12-4}$	.47	$r_{14-4}$	-.36
$r_{12-3}$	.24	$r_{24-2}$	-.13
$r_{12-1}$	.45	$r_{24-4}$	-.08

From these coefficients of the first order we may now compute those of the second order. The formula is similar in form to the preceding except that we are expressing a coefficient with two secondary subscripts in terms of coefficients with one secondary subscript.

$$r_{14-23} = \frac{r_{14-2} - r_{12-2} r_{24-2}}{\sqrt{1 - r_{12-2}^2} \sqrt{1 - r_{24-2}^2}}$$

The similarity of this formula to the preceding is obvious. The secondary subscript is the same throughout (2). The primary subscripts of the first term in the numerator are the same as the primary of the coefficient for which we are solving (14). The secondary subscript of the coefficient for which we are solving that does not appear as a secondary in the numerator (3) appears in both primaries in the second term of the numerator. The primaries (14) of the coefficient for which we are solving also appear as primaries in the second term in the numerator — one in each factor. The subscripts in the denominator are the same as those in the second term of the numerator. Substituting the proper values in this formula we have:

$$\begin{aligned} r_{14-23} &= \frac{-.34 - .24 \times (-.08)}{\sqrt{1 - .24^2} \sqrt{1 - (-.08)^2}} = \frac{-.34 - (-.019)}{\sqrt{.942} \sqrt{.994}} \\ &= \frac{-.34 + .019}{.971 \times .994} = \frac{-.321}{.968} = -.33 \end{aligned}$$

It is possible to compute this same coefficient by another formula as a check:

$$r_{14-23} = \frac{r_{14-3} - r_{12-3} r_{24-3}}{\sqrt{1 - r_{12-3}^2} \sqrt{1 - r_{24-3}^2}}$$

This conforms to the specifications mentioned in explaining the other formula for  $r_{14-23}$ , only it uses a different set of first order coefficients. In this case the secondary subscript that appears throughout is 3 instead of 2. Substituting:

$$r_{14-23} = \frac{-.36 - .28 \times (-.13)}{\sqrt{1 - .28^2} \sqrt{1 - (-.13)^2}} = \frac{-.36 - (-.036)}{\sqrt{.922} \sqrt{.983}}$$

$$= \frac{-.36 + .036}{.960 \times .991} = \frac{-.324}{.951} = -.34$$

This checks approximately with the result of the other formula. The difference of .01 is due to the fact that the coefficients of the first order were computed merely to two decimal places. If further decimals had been retained, the check would theoretically be perfect. If the proper coefficients of the first order are available, it is possible to compute all of those of the second order in two ways to detect any mistakes in the work up to that point. Making similar computation for the other coefficients that are necessary in the present problem we have:

$r_{13-24}$	.25
$r_{12-24}$	.23
$r_{14-22}$	-.33

Before we can compute the regression equation we need to know the average (or mean) of each variable as well as its standard deviation ( $\sigma$ ). These figures are as follows:

	Mean	$\sigma$
$X_1$	.00	.72
$X_2$	28	10
$X_3$	19	6
$X_4$	210	15

The notation  $X_1$  indicates original scores in the criterion and they were so arranged that their mean was 0. Their standard deviation was .72.  $X_2$  denotes score in test 2. Its average was 28 and its standard deviation 10, etc. The formula for the regression equation is:

$$x_1 = \frac{\sigma_{1-234}}{\sigma_{2-134}} r_{12-24} x_2 + \frac{\sigma_{1-234}}{\sigma_{3-124}} r_{13-24} x_3 + \frac{\sigma_{1-234}}{\sigma_{4-123}} r_{14-23}$$

The  $r$  factors are the partial correlation coefficients computed previously. The  $x$  values represent deviations of a particular measure from the mean of that measure,  $x_2$ , for instance, indicating the deviation of a measure from the mean score in test 2. The  $\sigma$  values, which represent the standard deviation of a variable with the effect of the others eliminated, must be computed thus:

$$\sigma_{1-234} = \sigma_1 \sqrt{1 - r_{12}^2} \sqrt{1 - r_{13-2}^2} \sqrt{1 - r_{14-23}^2}$$

$$\sigma_{2-134} = \sigma_2 \sqrt{1 - r_{23}^2} \sqrt{1 - r_{24-3}^2} \sqrt{1 - r_{12-24}^2}$$

$$\sigma_{3-124} = \sigma_3 \sqrt{1 - r_{32}^2} \sqrt{1 - r_{34-2}^2} \sqrt{1 - r_{12-24}^2}$$

$$\sigma_{4-123} = \sigma_4 \sqrt{1 - r_{42}^2} \sqrt{1 - r_{43-2}^2} \sqrt{1 - r_{12-24}^2}$$

The first factor in each product is the ordinary standard deviation of the variable whose number appears as the primary subscript at the left side of the equation. In  $\sigma_{1-234}$  the first subscript is of the zero order (12); the next one is of the first order and is obtained by putting the 2 into the secondary and bringing in another primary (3); for the last factor the 3 goes over into the secondary and the remaining one (4) is brought into the primary. It is to be noted that 1 appears as a primary throughout. The other formulæ embody the following principles: The subscript that appears as the primary at the left of the equation remains as a primary throughout. The last factor always has 1 as one of its primaries. Hence the last factor can always be determined by using as primaries 1 and the primary that appears at the left of the equation and using all the other variables as secondaries. The next to the last factor is obtained by shifting one of the secondaries into the primary, displacing the primary that is not to be a primary throughout. For instance, take  $\sigma_{4-123}$ . The last factor must have 1 as a primary subscript and also 4 which appears at the left of the equation as a primary. The next to the last factor drops 2 from the secondary and puts it in the primary to replace 1. It cannot replace 4 because that must remain as a primary throughout. The first factor now has this secondary 3 dropped and moved into the primary, replacing 2. Substituting the appropriate values in these equations:

$$\sigma_{1-234} = .72 \sqrt{1-.51^2} \sqrt{1-.24^2} \sqrt{1-(-.33)^2} = .72 \times .862 \times .971 \times .944 = .573$$

$$\sigma_{2-134} = 10 \sqrt{1-.66^2} \sqrt{1-(-.13)^2} \sqrt{1-.25^2} = 10 \times .751 \times .991 \times .968 = 7.15$$

$$\sigma_{3-124} = 6 \sqrt{1-.66^2} \sqrt{1-(-.08)^2} \sqrt{1-.23^2} = 6 \times .751 \times .997 \times .973 = 4.38$$

$$\sigma_{4-123} = 15 \sqrt{1-(-.22)^2} \sqrt{1-(-.13)^2} \sqrt{1-(-.33)^2} = 15 \times .975 \times .991 \times .944 = 13.68$$

We are now ready to substitute in the regression equation.

$$\begin{aligned} x_1 &= \frac{.573}{7.15} .25 x_2 + \frac{.573}{4.38} .23 x_3 - \frac{.573}{13.68} .33 x_4 \\ &= .02 x_2 + .03 x_3 - .014 x_4 \end{aligned}$$

There is one further step to take before the equation is in its most useful form. As above given, it involves the deviations of the scores from their mean rather than the actual original scores. If it were to be used in this form, it would be necessary to convert each measure into a deviation and substitute in the equation; then, after  $x_1$  had been obtained, to convert it back into terms of actual score. It is better to make a single transformation for the whole equation so that original scores can be substituted in it directly. This can be done by virtue of the fact that a deviation is simply the original score minus the mean, so that  $x_1 = X_1 - M_1$ ,  $x_2 = X_2 - M_2$ , etc., where  $x_1$  is the deviation of the criterion,  $X_1$  the original score, and  $M_1$  the mean of the criterion scores, and the same meaning is attached to  $x_2$ , the deviation of a score in test 2, etc. The mean scores have been given above so that we may make the following substitutions:



## APPENDIX

525

$$\begin{aligned}x_1 &= X_1 - 0 \\x_2 &= X_2 - 28 \\x_3 &= X_3 - 19 \\x_4 &= X_4 - 210\end{aligned}$$

Making this transformation we have:

$$\begin{aligned}X_1 - 0 &= .02 (X_2 - 28) + .03 (X_3 - 19) - .014 (X_4 - 210) \\X_1 &= .02 X_2 + .03 X_3 - .014 X_4 + 1.82.\end{aligned}$$

This is the final form of the regression equation, and if a given applicant has taken the three tests his scores may be substituted in this equation to obtain his most probable score in the criterion. (Cf. p. 238.)

When more than four variables are involved, the labor of computing the coefficients increases, but the procedure above outlined does not have to be followed, for there are various short cuts available. Even so the technique of partial correlation is tedious, but worth while.

No effort has been made in the foregoing to present the theory of partial correlation. This lies beyond the scope of the present work. For a discussion of the theory reference is made to Yule (692); for short cuts in the computation of partial coefficients, to Kelly (272); and for short cuts that obviate the use of certain coefficients, to Rosenow (490 — appendix).

# BIBLIOGRAPHY

## FOREWORD

THE bibliography does not include references that:

1. Deal with employment problems, but contain no psychology.
2. Deal with general psychology, but have no relation to employment.
3. Describe mental tests unless actually used for employment purposes — with the exception of a few tests that are especially novel.
4. Discuss vocational *guidance* unless vocational *selection* is at least indirectly involved.

It includes a few miscellaneous articles to which reference is made in the text, a few of historical interest, and a number on general statistics. The topic of job analysis has a large literature that almost borders on psychology and some typical references on this borderline are given.

The bibliography is not classified for the reason that so many of the references would have to appear several times. The reader interested in following up a certain topic will find that the chapter on that topic contains bracketed numbers referring to the outstanding titles in the bibliography dealing with that topic.

All available references have been examined personally and evaluated before inclusion. (A few that were not accessible had to be evaluated by the title alone.) In many instances a business article with a psychological title was found to contain no psychology or merely glittering generalities, and was discarded. The author has tried to include no references that would actually waste the time of a reader interested in employment psychology. The bibliography is up to date of February 1, 1926.

In referring to articles the form of citation used by the *Psychological Review* publications has been followed, namely, the year, then the volume in bold-face figures, and then the pages.

1. Achilles, P. S. and E. M. Estimates of the Military Value of Certain Character Qualities. *Journal of Applied Psychology*, 1917, **1**, 305-316.
2. Adams, H. F. The Mythology and Science of Character Analysis. *Scribner's*, 1921, **69**, 569-575.
3. Adams, H. F. Psychology Goldbricks. *Scribner's*, 1921, **70**, 94-101.
4. Allen, E. G. Adopting Standards to Meet Trade Training Requirements. *Annals of the American Academy of Political and Social Science*, 1916, **65**, 205-208.
5. Allen, F. J. A Guide to the Study of Occupations. Cambridge, Harvard University Press, 1921, pp. 183.

6. Allen, G. L. A Correlation Between Mental Tests and Manual Aptitudes. Columbia University, 1916 (Master's Essay).
7. Allport, F. H. and G. W. Personality Traits: Their Classification and Measurement. *Journal of Abnormal Psychology and Social Psychology*, 1921, 16, 6-40.
8. Allport, G. W. Personality and Character. *Psychological Bulletin*, 1921, 18, 441-455.
9. American Psychological Association. Report of the Committee on the Standardizing of Procedure in Experimental Tests. *Psychological Monographs*, 1910, 13, whole no. 53, pp. 108.
10. Anderson, L. D. Estimating Intelligence by Means of Printed Photographs. *Journal of Applied Psychology*, 1921, 5, 152-155.
11. Andrews, L. G. A Grading System for Picking Men. *Sales Management*, 1922, 4, 143-144.
12. Andrews, L. G. Scoring of Personal History Records. Bureau of Personnel Research, Carnegie Institute of Technology, Pittsburgh, Report A, 1921.
13. Anon. How the Army Uses Individual Differences in Experience. *Psychological Bulletin*, 1918, 15, 187-206.
14. Anon. Measurement and Utilization of Brain Power in the Army. *Science*, 1919, 49, 221-226, 251-259.
15. Anon. Psychotechnische Eignungsprüfung von Feuerwehrlenten in Dresden. *Feuerwehrtechnische Zeitschrift*, 1918, 6, 20.
16. Anon. Selecting Employees. New York, American Management Association, 1923. Committee report series no. 6, pp. 24.
17. Anon. Selecting the Supervisory Forces. New York, American Management Association, 1923, Committee report series no. 4, pp. 15.
18. Anon. Testing Men Like Material. *Factory*, 1923, 30, 294-295.
19. Anon. Trade Tests. *Personnel*, 1919, 1, no. 2, pp. 1, 3, 6.
20. Arnold, R. Beiträge zur Eignungsprüfung für Tischlergewerbe. *Zeitschrift für Angewandte Psychologie*, 1925, 25, 88-116.
21. Atkins, P. M. Employee Specifications. *Industrial Management*, 1924, 68, 115-118.
22. Ayres, L. P. Psychological Tests in Vocational Guidance. *Journal of Educational Psychology*, 1913, 4, 232-237.
23. Bagby, E. The Psychological Effects of Oxygen Deprivation. *Journal of Comparative Psychology*, 1921, 1, 97-113.
24. Bailey, W. G. A Method of Rating Employees for Promotion and Force Reduction. *Public Personnel Studies*, 1925, 3, 246-251.
25. Baumgarten, F. Beiträge zur Berufskunde des Versicherungswesens. II. Zur Psychotechnik und Charakterologie der Regulierungsbeamten. *Zeitschrift für Angewandte Psychologie*, 1925, 25, 1-64.
26. Baumgarten, F. Zur Psychologie und Psychotechnik des Ver-

- sicherungsagenten. *Zeitschrift für Angewandte Psychologie*, 1924, 23, 21-80.
27. Benary, W. Kurzer Bericht über Arbeiten zu Eignungsprüfungen für Flieger-Beobachter. *Zeitschrift für Angewandte Psychologie*, 1920, 16, 250-308.
28. Bengé, E. J. Analyzing the Employee and His Job. *Factory*, 1923, 30, 424-426, 468.
29. Bengé, E. J. Are Rating Scales Justified? *Industrial Management*, 1923, 65, 192-202.
30. Bengé, E. J. Building a Better Force of Workers; Four Trade Tests. *Factory*, 1920, 25, 1237-1239.
31. Bengé, E. J. Proficiency Tests for Stenographers. *Management and Administration*, 1924, 8, 615-616.
32. Bengé, E. J. Selecting the Salaried Employee. *Industrial Management*, 1925, 69, 279-283.
33. Bengé, E. J. Selection Tests. The Atlantic Refining Co. *American Management Review*, 1923, 12, no. 5, p. 6.
34. Bengé, E. J. Simple Tests for Selecting Office Workers. *Industrial Management*, 1921, 61, 91-93.
35. Bengé, E. J. Standard Practice in Personnel Work. New York, Wilson, 1920, pp. 286.
36. Bengé, E. J. and McTague, L. L. The Value of Ratings. *Industrial Management*, 1922, 64, 239-242.
37. Berliner, A. Einfluss des Alters und die Geschwindigkeit bei leichter Arbeit. *Zeitschrift für Angewandte Psychologie*, 1924, 23, 196-204.
38. Berry, W. S. Psychological Testing. Scovill Mfg. Co. *American Management Review*, 1923, 12, no. 6, pp. 9-10.
39. Berry, W. S. Testing Tests. *Scovill Bulletin*, 1923, 9, 5-7.
40. Bieneman, D. Recherches sur l'aptitude dactylographique en vue de l'orientation professionnelle. *Bureau internationale du travail études et doc.*, Series J, Genève, 1923, pp. 48.
41. Bills, M. A. A Method for Classifying the Jobs and Rating the Efficiency of the Clerical Workers. *Journal of Personnel Research*, 1923, 1, 384-393.
42. Bills, M. A. Methods for the Selection of Comptometer Operators and Stenographers. *Journal of Applied Psychology*, 1921, 5, 275-283.
43. Bills, M. A. Predicting the Careers of Clerical Workers. *Forbes*, 1923, 13, 91-93; 103-105.
44. Bills, M. A. Predicting Managerial Success: A Case Study of Two Business Men. *Journal of Personnel Research*, 1925, 4, 46-51.
45. Bills, M. A. Relation of Mental Alertness Test Score to Position and Permanency in Company. *Journal of Applied Psychology*, 1923, 7, 154-156.

46. Bills, M. A. Social Status of the Clerical Worker and His Permanence on the Job. *Journal of Applied Psychology*, 1925, 9, 424-427.
47. Bills, M. A. Standardizing the Selection of Clerical Workers. *Service Bulletin of the Bureau of Personnel Research*, Carnegie Institute of Technology, 1923, 5, no. 9.
48. Bills, M. A. A Test for Use in the Selection of Stenographers. *Journal of Applied Psychology*, 1921, 5, 373-377.
49. Binet, A., and Simon, T. Le développement de l'intelligence chez les enfants. *L'Année psychologique*, 1908, 14, 1-94.
50. Bingham, W. V. Applied Psychology at Carnegie Institute of Technology. *Annual Report of the Director, Carnegie Institute of Technology*, 1917, 14, 53-60.
51. Bingham, W. V. Army Personnel Work. *Journal of Applied Psychology*, 1919, 3, 1-12.
52. Bingham, W. V. Coöperative Business Research. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 179-189.
53. Bingham, W. V. Development and Standardization of Army Trade Tests. *Industrial Education Circular*, Bureau of Education, Washington, 1919, 4, 11-16.
54. Bingham, W. V. The Division of Coöperative Research. *Carnegie Technical Journal*, 1921, 2, 12-13.
55. Bingham, W. V. Intelligence Test Scores and Business Success. *Psychological Bulletin*, 1924, 21, 103-105.
56. Bingham, W. V. Measuring a Workman's Skill. National Society for Vocational Education, Bulletin no. 30, Proceedings of St. Louis Convention, 1919, 7-17.
57. Bingham, W. V. Methods of Research in Industrial Relations. *Journal of Personnel Research*, 1925, 4, 127-132.
58. Bingham, W. V. On the Possibility of an Applied Psychology. *Psychological Review*, 1923, 30, 289-305.
59. Bingham, W. V. Psychology Applied. *Scientific Monthly*, 1923, 16, 141-159.
60. Bingham, W. V. Standardized Mental Test in Personnel Selection. *Journal of Society of Automotive Engineers*, 1923, 12, 272.
61. Bingham, W. V. Student Personnel Service and Industrial Research. *Journal of Personnel Research*, 1923, 2, 55-64.
62. Bingham, W. V. What Industrial Psychology Asks of Management. *Bulletin of the Taylor Society*, 1924, 9, 243-248.
63. Bingham, W. V. The Work of the Personnel Research Federation. *Journal of Personnel Research*, 1926, 4, 383-390.
64. Bingham, W. V., and Davis, W. T. Intelligence Test Scores and Business Success. *Journal of Applied Psychology*, 1924, 8, 1-22.

65. Bixler, H. H. Ten Minute Intelligence Test in Junior Employment Offices. *School and Society*, 1921, 14, 166-168.
66. Bloomfield, D. (Editor). *Employment Management*. New York, Wilson, 1919, pp. 507.
67. Bloomfield, D. (Editor). *Problems in Personnel Management*. New York, Wilson, 1923, pp. 557.
68. Blumenfeld, W. Eignungsprüfungen an Lehrlingen der Metall-industrie im Psychotechnischen Institut und der T. H. zu Dresden. *Maschinenbau*, 1923, 2, 249-253.
69. Bogen, H. Beiträge zur kunstgewerblichen Eignungsprüfung. *Industrielle Psychotechnick*, 1925, 2, 12-21.
70. Bogen, H. Die Psychologische Abteilung in der Praxis des Berufs-amtes. *Praktische Psychologie*, 1923, 4, 303-308.
71. Book, W. F. Voluntary Motor Ability of the World's Champion Typists. *Journal of Applied Psychology*, 1924, 8, 283-308.
72. Borremans, Ponthière, et al. L'orientation professionnelle. Bruxelles, 1923, pp. 470.
73. Brandenburg, G. C. Personality and Vocational Achievement. *Journal of Applied Psychology*, 1925, 9, 281-292.
74. Braunhausen, N. Psychologische Personalbogen als Helfmittel für Pädagogik und Berufsberatung. *Zeitschrift für Angewandte Psychologie*, 1919, 15, 1-32.
75. Bregman, E. O. Studies in Industrial Psychology. *Archives of Psychology*, 1922, 9, whole no. 59, pp. 60.
76. Bregman, E. O. A Study in Industrial Psychology — Tests for Special Abilities. *Journal of Applied Psychology*, 1921, 5, 127-151.
77. Brewer, J. M., and Kelly, R. W. A Selected Critical Bibliography of Vocational Guidance. Cambridge, Harvard University Press, 1917, Harvard Bulletins in Education, no. 4, pp. 76.
78. Brewington, A. Prognostic Tests for Typewriting. *American Shorthand Teacher*, 1923, 4, nos. 1 and 2.
79. Bridges, J. W., and Coler, L. E. The Relation of Intelligence to Social Status. *Psychological Review*, 1917, 24, 1-31.
80. Bridges, J. W., and Dollinger, M. The Correlation Between Interests and Abilities in College Courses. *Psychological Review*, 1920, 27, 308-314.
81. Brigham, C. C. A Study of American Intelligence. Princeton, Princeton University Press, 1923, pp. 210.
82. Bronner, A. F. Psychology of Special Abilities and Disabilities. Boston, Little, Brown & Company, 1917, pp. 269.
83. Brotemarkle, R. A. A Comparison Test for Investigating the Ideational Content of the Moral Concepts. *Journal of Applied Psychology*, 1922, 6, 235-242.
84. Brown, W. M. Character Traits as Factors in Intelligence Test Performance. *Archives of Psychology*, 1924, no. 65, pp. 66.

85. Brown, W. M. A Study of the "Caution" Factor and its Importance in Intelligence Test Performance. *American Journal of Psychology*, 1924, 35, 368-386.
86. Brown, W. M. A Study of the Predictive Value of Certain Kinds of Scores in Intelligence Tests. *Journal of Educational Psychology*, 1924, 15, 448-461.
87. Brubacher, A. R. The Teaching Personality Quotient. *Journal of Applied Psychology*, 1921, 5, 39-45.
88. Bureau of Personnel Research. An Accountancy Test. Carnegie Institute of Technology, Pittsburgh, 1918.
89. Bureau of Public Personnel Administration. A Method of Rating the History and Achievements of Applicants for Positions. *Public Personnel Studies*, 1925, 3, 202-209.
90. Bureau of Public Personnel Administration. Methods of Selecting Employees to Fill High Grade Positions in the Public Service. *Public Personnel Studies*, 1924, 2, 13-58.
91. Bureau of Public Personnel Administration. Partially Standardized Tests for Junior Clerk. *Public Personnel Studies*, 1925, 3, 346-371.
92. Bureau of Public Personnel Administration. Suggested Tests for Female Playground Supervisor. *Public Personnel Studies*, 1925, 3, 225-236.
93. Bureau of Public Personnel Administration. Suggested Tests for Plumber. *Public Personnel Studies*, 1925, 3, 289-300.
94. Bureau of Public Personnel Administration. Suggested Tests for Senior Library Assistant, Circulation Department. *Public Personnel Studies*, 1925, 3, 97-109.
95. Burke, R. J. Written Specifications for Hiring. *Annals of the American Academy of Political and Social Science*, 1916, 65, 176-181.
96. Burnet, I. An Experimental Investigation of Repetitive Work. *Journal of the National Institute of Industrial Psychology*, 1924, 2, 18-23.
97. Burr, E. T. Activities of the Vocational Adjustment Bureau for Girls. *National Vocational Guidance Association Bulletin*, 1924, 2, no. 5.
98. Burr, E. T. Minimum Intellectual Levels of Accomplishment in Industry. *Journal of Personnel Research*, 1924, 3, 207-212.
99. Burr, E. T. Psychological Tests Applied to Factory Workers. *Archives of Psychology*, 1922, whole no. 55, pp. 91.
100. Burt, C. Tests for Clerical Occupations. *Journal of the National Institute of Industrial Psychology*, 1922, 1, 23-27, 79-81.
101. Burt, C. Vocational Diagnosis in Industry and at School. In *Lectures on Industrial Administration* (edited by Muscio). London, Routledge, 1920, 79-120.

102. Burt, H. E. Employment Psychology in the Rubber Industry. *Journal of Applied Psychology*, 1920, 4, 1-17.
103. Burt, H. E. Measuring Interests Objectively. *School and Society*, 1923, 17, 444-448.
104. Burt, H. E. Professor Münsterberg's Vocational Tests. *Journal of Applied Psychology*, 1917, 1, 201-213.
105. Burt, H. E., and Ives, F. Vocational Tests for Agricultural Engineers. *Journal of Applied Psychology*, 1923, 7, 178-187.
106. Busse, Dr. Berufskundliche Untersuchung des Rangierdienstes. *Praktische Psychologie*, 1923, 4, 350-355.
107. Cades, E. The Textile Industry in Philadelphia. *Psychological Clinic*, 1924, 15, 203-228.
108. Cady, V. M. The Estimation of Juvenile Incurability. *Journal of Delinquency*, Monograph no. 2, California Bureau of Juvenile Research, Whittier State School, 1923, pp. 140.
109. Campbell, B. P. Tests for Shop and Office Jobs. *Management and Administration*, 1924, 8, 409-412.
110. Camus, J., and Nepper. Recherches sur l'aptitude à l'aviation. *Bulletin de l'Institut générale Psychologique*, 1917, 1-3.
111. Carnegie Institute of Technology. Aids in Selecting Salesmen. Pittsburgh, Bureau of Salesmanship Research, 1917. (Published privately.)
112. Carney, C. S. Some Experiments with Mental Tests as an Aid in the Selection and Placement of Clerical Workers in a Large Factory. *Conference on Educational Measurements*, Indiana University, 1919, 5, 60-74.
113. Cattell, J. McK. The Psychological Corporation. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 165-171.
114. Cattell, J. McK., and Farrand, L. Physical and Mental Measurements of the Students of Columbia University. *Psychological Review*, 1896, 3, 618-648.
115. Cauer, M. Berufspsychologische und psychotechnische Studie zur Eignungsprüfung der Schneiderin. *Industrielle Psychotechnik*, 1924, 1, 73-82.
116. Chapman, J. C. The Individual Injustice and Guessing in the True False Examination. *Journal of Applied Psychology*, 1922, 6, 342-348.
117. Chapman, J. C. Mental Tests in Industry. *Personnel*, 1919, 1, no. 3, pp. 1 and 9.
118. Chapman, J. C. Tests for Trade Proficiency. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 45-59.
119. Chapman, J. C. Trade Tests. New York, Holt, 1921, pp. 435.
120. Chapman, J. C., and Toops, H. A. A Written Trade Test. *Journal of Applied Psychology*, 1919, 3, 358-365.



121. Charters, W. W. Activity Analysis and Curriculum Construction. *Journal of Educational Research*, 1922, 5, 357-367.
122. Charters, W. W. The Collecting of Unrecorded Specifics. *Journal of Educational Research*, 1922, 5, 280-294.
123. Charters, W. W., and Whitley, I. B. Analysis of Secretarial Duties and Traits. Baltimore, Warwick and York, 1924, pp. 186.
124. Christiaens, A. G. L'orientation professionnelle en Belgique. *Notes et Mémoires de l'Institut Lannelongue*, 1923, 4, 89-131.
125. Claparède, E. Problems and Methods of Vocational Guidance. Studies and Reports from the International Labor Office, 1922, Series J, no. 1, pp. 71.
126. Clark, J. C. Job Analysis and Job Specification. *Railway Age*, 1921, 71, 445-446.
127. Clary, H. L. The Zoning of Jobs. *Industrial Management*, 1921, 61, 324-329.
128. Cleeton, G. U., and Knight, F. B. Validity of Character Judgments Based on External Criteria. *Journal of Applied Psychology*, 1924, 8, 215-231.
129. Clothier, R. C. Organization for an Occupational Survey. *Journal of Personnel Research*, 1923, 1, 427-450.
130. Clothier, R. C. Selective Functions of the Employment Department. *Efficiency Society Journal*, 1916, 5, 237-248.
131. Cody, S. Commercial Tests and How to Use them. Yonkers, World Book Company, 1918, pp. 216.
132. Cogan, L. C., Conklin, A. M., and Hollingworth, H. L. An Experimental Study of Self-Analysis, Estimates of Associates and the Results of Tests. *School and Society*, 1915, 2, 171-179.
133. Coover, J. E. Experiments in Psychological Research, Stanford University Publications, 1917, pp. 641.
134. Coss, J. J., and Outhwaite, L. Personnel Management. Topical Outline and Bibliography. Washington, Office of the Adjutant General, 1919, pp. 59.
135. Couvé, R. Bericht über die psychotechnische Untersuchung von zwei Eisenbahnszusammenstößen. *Industrielle Psychotechnik*, 1924, 1, 160-168.
136. Couvé, R. Die Psychotechnische Prüfung von Eisenbahnverkehrsbeamten. *Industrielle Psychotechnik*, 1924, 1, 22-29.
137. Cowdery, K. M. Measures of General Intelligence as Indices of Success in Trade Learning. *Journal of Applied Psychology*, 1922, 6, 311-330.
138. Craig, D. R. The Preference Interest Questionnaire in Selecting Retail Saleswomen. *Journal of Personnel Research*, 1925, 3, 366-375.
139. Dennison Manufacturing Company. Selection of Salesmen. *American Management Review*, 1923, 12, no. 7, pp. 12.

## BIBLIOGRAPHY

535

140. Dietz, J. W. Some Aspects of Personnel Research in a Manufacturing Organization. *Annals of the American Academy of Political and Social Science*, 1925, 119, no. 208, 103-107.
141. Dockeray, F. C., and Isaacs, S. Psychological Research in Aviation in Italy, France, England, and the American Expeditionary Forces. *Journal of Comparative Psychology*, 1921, 1, 115-148.
142. Dodge, R. Mental Engineering During the War. *American Review of Reviews*, 1919, 59, 504-508.
143. Dodge, R. The Psychology of Auto Drivers. *Science*, 1922, n.s. 56, Supplement Dec. 29, p. ix.
144. Dolar, H. H. Developments in the Field of Mental Testing. *Public Health Reports*, 1924, 39, 2505-2552.
145. Dooley, C. Choosing the Office Personnel. *Engineering and Industrial Management*, 1921, 5, 125-126.
146. Douglas, A. A. Vocational Interests of High School Seniors. *School and Society*, 1922, 16, 79-84.
147. Downey, J. E. Testing the Will Temperament Tests. *School and Society*, 1922, 16, 161-168.
148. Downey, J. E. The Will Temperament and Its Testing. Yonkers, World Book Company, 1922, pp. 399.
149. Drever, J. The Psychology of Industry. New York, Dutton, 1921, pp. 148.
150. Dück, J. Die Berufseignung der Kanzleiangestellten. *Zeitschrift für Angewandte Psychologie*, 1919, 14, 285-305.
151. Dunlap, K. Fact and Fable in Character Analysis. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 74-80.
152. Dunlap, K. Psychological Research in Aviation. *Science*, 1919, 49, 94-97.
153. Dunlap, K. The Reading of Character from External Signs. *Scientific Monthly*, 1922, 15, 153-165.
154. Dunn, W. E. Measuring Executive Ability. *Industrial Management*, 1922, 63, 292-295.
155. Dutton, W. S. Fewer Mistakes in Selecting Employees. *Factory*, 1921, 26, 38-39.
156. Earle, F. M. Vocational Guidance and Selection. *Journal of the National Institute of Industrial Psychology*, 1925, 2, 346-351.
157. Edie, L. D. Practical Psychology for Business Executives. New York, Wilson, 1922, pp. 392.
158. Ewer, B. C. Applied Psychology. New York, Macmillan, 1923, pp. 480.
159. Ewer, B. C. First Step in the Choice of Salesmen. *Printers' Ink*, 1923, 124, 3-4, 6, 157-158.
160. Federal Board for Vocational Education. Bibliography of Employment Management. Bulletin 57, Washington, 1920.

161. Federal Board for Vocational Education. Bibliography of Vocational Guidance. Bulletin 66, Washington, 1921.
162. Feingold, C. A. The Relation between the Intelligence and Vocational Choices of High School Pupils. *Journal of Applied Psychology*, 1923, 7, 143-153.
163. Feiss, R. A. Personal Relationship as a Basis of Scientific Management. *Annals of the American Academy of Political and Social Science*, 1916, 65, 27-56.
164. Fernald, G. M., and Sullivan, E. B. Intelligence Tests of the Los Angeles Police Force. *Psychological Bulletin*, 1925, 22, 335-336.
165. Ferree, C. E., and Rand, G. A Study of Ocular Functions with Special Reference to the Look-out and Signal Service of the Navy. *Psychological Bulletin*, 1920, 17, 77-78.
166. Filer, H. A., and O'Rourke, L. J. Progress in Civil Service Tests. *Journal of Personnel Research*, 1923, 1, 484-520.
167. Fish, E. H. How Shall We Select Our Executives? *Management and Administration*, 1923, 6, 183-184.
168. Fish, E. H. Human Engineering. *Journal of Applied Psychology*, 1917, 1, 161-174.
169. Flanders, J. K. Mental Tests of a Group of Employed Men Showing Correlations with Estimates Furnished by Employers. *Journal of Applied Psychology*, 1918, 2, 197-206.
170. Fontègne, U. J., and Solari, E. Le Travail de Téléphoniste. *Archives de Psychologie*, 1918, 66, 81-136.
171. Foster, W. R. Vocational Selection in a Chocolate Factory. *Journal of the National Institute of Industrial Psychology*, 1924, 2, 159-163.
172. Frankel, L. K., and Fleisher, A. The Human Factor in Industry. New York, Macmillan, 1920, pp. 366.
173. Fraser, J. A. The Vocational Selection and Training of Operatives for the Weaving Industry. *Journal of the National Institute of Industrial Psychology*, 1924, 2, 88-93.
174. Freyd, M. The Graphic Rating Scale. *Journal of Educational Psychology*, 1923, 14, 83-102.
175. Freyd, M. A Graphic Rating Scale for Teachers. *Journal of Educational Research*, 1923, 8, 433-439.
176. Freyd, M. Measurement in Vocational Selection. *Journal of Personnel Research*, 1923, 2, 215-249, 268-284, 377-385.
177. Freyd, M. The Measurement of Interests in Vocational Selection. *Journal of Personnel Research*, 1922, 1, 319-328.
178. Freyd, M. A Method for the Study of Vocational Interests. *Journal of Applied Psychology*, 1922, 6, 243-254.
179. Freyd, M. Personalities of the Socially and the Mechanically Inclined. *Psychological Monographs*, 1923, 33, whole no. 151, pp. 191.

180. Freyd, M. The Statistical Viewpoint in Vocational Selection. *Journal of Applied Psychology*, 1925, 9, 349-356.
181. Freyd, M. Test Series for Journalistic Aptitude. *Journal of Applied Psychology*, 1921, 5, 46-56.
182. Friedman, H. Eignungsprüfung zum Setzer und Druckerberufe in der Buchdruckerlehranstalt zu Leipzig. *Praktische Psychologie*, 1920, 1, 249-252.
183. Friedrich, Adolph. Die Analyse des Schlosserberufs. *Praktische Psychologie*, 1922, 3, 287-300.
184. Friedrich, Adolph. Die Einstellungsprüfung der Schlosser und Dreherlinge innerhalb der Friedr. Krupp A. G. Essen. *Praktische Psychologie*, 1922, 3, 159-166.
185. Frois, M. L'Organisation technique du travail et l'orientation professionnelle. *Notes et Mémoires, l'Institut Lannelongue*, 1923, 41, 76-89.
186. Fryer, D. General Knowledge Tests. *The Engineer*, 1922, 133, 45-46.
187. Fryer, D. Getting Personality into the Plant. *Factory*, 1921, 27, 46.
188. Fryer, D. Industrial Dissatisfaction. *Industrial Psychology*, 1926, 1, 25-29.
189. Fryer, D. Intelligence and Interest in Vocational Adjustment. *Pedagogical Seminary*, 1923, 30, 127-151.
190. Fryer, D. Occupational Intelligence Standards. *School and Society*, 1922, 16, 273-277.
191. Fryer, D. Psychology and Industry in France and Great Britain. *Journal of Personnel Research*, 1924, 2, 396-402.
192. Fryer, D. The Significance of Interest for Vocational Progress. *Mental Hygiene*, 1924, 8, 466-505.
193. Fryer, D. Types of Work. A Note on Vocational Interest. *Journal of Applied Psychology*, 1925, 9, 281-292.
194. Fryer, D. Vocational Self-Guidance. Philadelphia, Lippincott, 1925, pp. 385.
195. Gambrill, B. L. College Achievement and Vocational Efficiency. New York, Teachers College Publications, 1922.
196. Garrett, H. E. An Empirical Study of Various Methods of Combining Incomplete Order of Merit Ratings. *Journal of Educational Psychology*, 1924, 15, 157-171.
197. Gaw, F. The Use of Performance Tests and Mechanical Tests in Vocational Guidance. *Journal of the National Institute of Industrial Psychology*, 1923, 1, 333-337.
198. Gerhardt, P. W. Psychological Tests for Workmen. *Industrial Management*, 1916, 51, 605-606.
199. Giese, F. Berufspsychologische Beobachtungen in Reichstele-

- graphen-Dienst. *Schriften zur Psychologie der Berufseignung und des Wirtschaftslebens*, 1923, 24, 1-74.
200. Giese, F. Zur Betriebsführung psychotechnischer Prüfstellen. *Praktische Psychologie*, 1921, 3, 1-12.
201. Gläsel, I. Von der Dresdner Prüfstelle der Reichsbahn. *Praktische Psychologie*, 1923, 4, 335-350.
202. Gley, E. Les Bases scientifiques de l'orientation professionnelle. *Notes et Mémoires l'Institut Lannelongue*, 1923, 4, 9-23.
203. Goddard, H. H. Human Efficiency and Levels of Intelligence. Princeton, Princeton University Press, 1920, pp. 128.
204. Goldman, H. The Applied Psychology of Hugo Münsterberg. *Journal of Applied Psychology*, 1918, 2, 116-129.
205. Goldsmith, D. Predicting the Salesman's Future Success by his Record During the First Year. Bureau of Personnel Research, Carnegie Institute of Technology, Report G, 1920.
206. Goldsmith, D. A Study of Five Hundred Personal History Blanks. New York, Guardian Life Insurance Company of America, 1922.
207. Goldsmith, D. The Use of the Personal History Blank as a Salesmanship Test. *Journal of Applied Psychology*, 1922, 6, 148-155.
208. Gould, R. L. Initial Twelve Month Production as an Index of Subsequent Value. *Proceedings of Fourth Annual Meeting, Bureau of Personnel Research*, Carnegie Institute of Technology, 1920.
209. Gowin, E. B. The Selection and Training of the Business Executive. New York, Macmillan, 1918, pp. 225.
210. Gradenwitz, A. Psychological Tests for Motormen. *Electric Railway Journal*, 1922, 59, 143-146.
211. Gradenwitz, A. Psychotechnics in Germany. *American Machinist*, 1920, 53, 407-408.
212. Greenwold, G. W. Simple Tests for Office Applicants. *Industrial Management*, 1919, 57, 377-378.
213. Griffenhagen, E. O. The Principles and Technique of Preparing an Occupational Classification of Positions in the Public Service. *Public Personnel Studies*, 1924, 2, 240-254.
214. Griffiths, C. H. Fundamentals of Vocational Psychology. New York, Macmillan, 1924, pp. 372.
215. Grimshaw, R. Performance Test Questions in Hiring Employees. *Machinery*, 1922, 29, 141-142.
216. Gronert, M. L. A Prognostic Test in Typewriting. *Journal of Educational Psychology*, 1925, 16, 182-185.
217. Gwinn, Thelma. The Value of Individual Tests in Predicting Success in an Engineering College. Unpublished Thesis, Carnegie Institute of Technology, 1923.
218. Hallbauer, U. Prüfung der Aufmerksamkeits und Reaktionsweise von Triebwagenführern. *Praktische Psychologie*, 1923, 4, 104-113.

219. Hamilton, A. M. Tests for the Perception of Efficiency of Recruits. *Medical Record*, 1918, 93, 285-286.
220. Hansen, C. F. Blanks Useful in Selecting Agents. *Service Bulletin, Bureau of Personnel Research*, Carnegie Institute of Technology, 1922, 4, pp. 11.
221. Hansen, C. F. Preliminary Study of the Ratings Given by Life Insurance School Students. *Office Report*, no. 5, *Bureau of Personnel Research*, Carnegie Institute of Technology, Sept., 1920, pp. 5.
222. Hansen, C. F. Psychology in the Service of the Life Insurance Business. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 190-193.
223. Hartman, R., and Dashiell, J. F. An Experiment to Determine the Relation of Interests to Abilities. *Psychological Bulletin*, 1919, 16, 259-262.
224. Hatt, E., and Brandenburg, G. C. Proposed Tests for Mechanical Ability. Society for the Promotion of Engineering Education, *Bulletin*, 1921, 12, 114-127.
225. Haight, B. F. A Scheme for Combining Incomplete Rankings. *Journal of Applied Psychology*, 1923, 7, 168-172.
226. Hayes, J. W. *Judging Human Character*, New York, Appleton, 1922.
227. Hayes, M. H. S., and Paterson, D. G. Experimental Development of the Graphic Rating Method. *Psychological Bulletin*, 1921, 18, 98-99.
228. Heilandt, A. Psychotechnische Eignungsprüfungen bei der Einstellung gewerblicher Lehrlings in der Werkschule der A. E. G. Fabriken Brunnenstrasse. *Betrieb*, 1920, 3, 16-21.
229. Heller, O. Eignungsprüfung und Unfallvorbeugung in der Holzindustrie. *Industrielle Psychotechnik*, 1924, 1, 99-118.
230. Henderschott, F. C. and Weakly, F. E. *The Employment Department and Employee Relations*. Chicago, La Salle Extension University, 1919, pp. 60.
231. Henmon, V. A. C. Air Service Tests of Aptitude for Flying. *Journal of Applied Psychology*, 1919, 3, 103-109.
232. Hepner, H. W. Better Judgments of Men. *Industrial Psychology*, 1926, 1, 19-24.
233. Herwig, B. Auswertungsverfahren bei der psychotechnischen Eignungsprüfung. *Praktische Psychologie*, 1920, 2, 45-48.
234. Heydt, C. Eignungsprüfungen für den Rangierdienst. *Industrielle Psychotechnik*, 1924, 1, 140-147.
235. Hildebrandt, H. Versuche mit psychotechnischen Eignungsprüfungen für Ingenieurlehrlinge. *Industrielle Psychotechnik*, 1925, 2, 42-49.
236. Hines, M., and O'Connor, J. A Measure of Finger Dexterity. *Journal of Personnel Research*, 1926, 4, 379-382.

237. Hirsch, A. Einstellung und Ausbildung Kaufmannischer Lehrlinge bei der Osranggesellschaft. *Praktische Psychologie*, 1921, 2, 329-332.
238. Hoke, E. R. The Measurement of Achievement in Shorthand. Johns Hopkins Studies in Education, no. 6, Johns Hopkins Press, 1922, pp. 118.
239. Hoke, E. R. Prognostic Tests of Stenographic Ability. New York, Gregg Publishing Co.
240. Holcombe, J. M., Jr. A Case of Sales Research. Report on First Steps in a Study of the Selection of Life Insurance Salesmen. *Bulletin of the Taylor Society*, 1922, 7, 112-121.
241. Hollingworth, H. L. Can Workers Be Tested? *Business Personnel*, 1920, Nov., 16-18.
242. Hollingworth, H. L. Correlation of Abilities as Affected by Practice. *Journal of Educational Psychology*, 1913, 4, 405-414.
243. Hollingworth, H. L. Individual Differences Before, During, and After Practice. *Psychological Review*, 1914, 21, 1-8.
244. Hollingworth, H. L. Judging Human Character. New York, Appleton, 1923, pp. 268.
245. Hollingworth, H. L. Specialized Vocational Tests and Methods. *School and Society*, 1915, 1, 918-922.
246. Hollingworth, H. L. Vocational Psychology, New York, Appleton, 1916, pp. 308.
247. Hollingworth, H. L., and Poffenberger, A. T. Applied Psychology. New York, Appleton, 1923, pp. 431.
248. Hollingworth, L. S. Comparison of the Sexes in Mental Traits. *Psychological Bulletin*, 1919, 16, 371-373.
249. Hooker, R. H. The Correlation of the Weather and the Crops. *Journal of the Royal Statistical Society*, 1907, 70, 1-42.
250. Hoopingarner, N. L. Psychological Tests Must be Based on Facts. *Electric Railway Journal*, 1922, 59, 716-717.
251. Hubbard, R. M. Interests Studied Quantitatively. *Journal of Personnel Research*, 1926, 4, 365-378.
252. Hubbell, W. D. Written Standard Job Specifications. *Industrial Management*, 1917, 54, 431-436.
253. Hull, C. L., and Limp, C. E. The Differentiation of the Aptitudes of an Individual by Means of Test Batteries. *Journal of Educational Psychology*, 1925, 16, 73-88.
254. Hull, C. L., and Montgomery, R. B. An Experimental Investigation of Certain Alleged Relations Between Character and Handwriting. *Psychological Review*, 1919, 26, 63-74.
255. Hunter, O. B., Cajigas, T., and Moss, F. A. Standardized Tests in Pathology. *Public Personnel Studies*, 1925, 3, 172-185.
256. Hunter, O. R., and Moss, F. A. Standardized Tests in Bacteriology. *Public Personnel Studies*, 1925, 3, 52-66.

257. Huth, A. Die Münchener Eignungsprüfung für Buchdrucker und Schriftsetzer. *Zeitschrift für Angewandte Psychologie*, 1922, 20, 365-390.
258. Ikin, A. G. The Qualities Desirable in a Foreman. *Journal of the National Institute of Industrial Psychology*, 1924, 2, 13-17.
259. Immig, G. Die Arbeitsprobe. *Praktische Psychologie*, 1921, 2, 338-344.
260. Immig, G. Psychotechnische Eignungsprüfungen in der Industrie. Die Eignungsprüfung für Lehrlinge bei der Firm Carl Zeiss, Jena. *Praktische Psychologie*, 1921, 2, 225-231.
261. Industrial Relations Association of America. *Proceedings*, May, 1920. Subject Sessions on: Employment Office Methods (pp. 231-262), Job Specifications (pp. 199-211), Mental Tests (pp. 271-280), Periodic Rating (pp. 311-327).
262. Jackson, W. M. Employee Tests. *Administration*, 1922, 4, 689-696.
263. Jaques, M. P. Mental Tests for Typists and Stenographers. *Industrial Management*, 1919, 58, 145-147.
264. Jones, E. S. Personality Terms Commonly Used in Recommendations. *Journal of Personnel Research*, 1924, 2, 421-430.
265. Jones, E. S. The Woolley Test Series Applied to the Detection of Ability in Telegraphy. *Journal of Educational Psychology*, 1917, 8, 27-34.
266. Jones, M. M. Classification of Personnel in the Army. In *Proceedings of Employment Managers' Conference*, at Rochester, May, 1918, Bulletin no. 247, U.S. Bureau of Labor Statistics, Washington, 1918, pp. 112-118.
267. Kafka, G. Zur Psychotechnik des Bremsen bei der elektrischen Strassenbahn. *Zeitschrift für Angewandte Psychologie*, 1921, Beiheft, 29, 102-106.
268. Kelly, R. W. Dangers in Rating Employees. *Industrial Management*, 1919, 58, 35-42.
269. Kelly, R. W. Hiring the Worker. New York, Engineering Magazine Company, 1918, pp. 250.
270. Kelly, R. W. Selecting and Training Interviewers. *Industrial Management*, 1919, 57, 263-270.
271. Kelly, T. L. Principles Underlying the Classification of Men. *Journal of Applied Psychology*, 1919, 3, 50-67.
272. Kelly, T. L. Statistical Method. New York, Macmillan, 1923, pp. 385.
273. Kemble, W. F. Choosing Employees by Test. New York, Engineering Magazine Company, 1917, pp. 333.
274. Kenagy, H. G. Characteristics of Successful Salesmen. *Service Bulletin, Bureau of Personnel Research*, Carnegie Institute of Technology, 1923, 5, no. 5, pp. 31.



275. Kenagy, H. G. Developments in Scientific Methods for Selecting Salesmen. *Sales Management*, 1924, 7, 1453 ff.
276. Kenagy, H. G. Empty Heads Make Poor Clerks. *Forbes*, 1922, 10, 180-181.
277. Kenagy, H. G. Some Characteristics of Successful Salesmen. *Printers' Ink*, 1923, 122, 33-48.
278. Kenagy, H. G., and Yoakum, C. E. The Selection and Training of Salesmen. New York, McGraw, 1925, pp. 380.
279. Kingsbury, F. A. Analyzing Ratings and Training Raters. *Journal of Personnel Research*, 1922, 1, 377-383.
280. Kingsbury, F. A. Applying Psychology to Business. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 2-12.
281. Kingsbury, F. A. Grading the Office Job. *Administration*, 1923, 5, 267-275, 393-401.
282. Kingsbury, F. A. Making Ratings Scales Work. *Journal of Personnel Research*, 1925, 4, 1-6.
283. Kingsbury, F. A. The Principles Involved in Securing Service Ratings as Exemplified in a Large Bank. *Public Personnel Studies*, 1925, 3, 70-84.
284. Kitson, H. D. A Critical Age as a Factor in Labor Turnover. *Journal of Industrial Hygiene*, 1922, 4, 199-202.
285. Kitson, H. D. Height and Weight as Factors in Salesmanship. *Journal of Personnel Research*, 1922, 1, 289-294.
286. Kitson, H. D. Industrial Psychology in Europe. *Journal of Applied Psychology*, 1921, 5, 287-290.
287. Kitson, H. D. Interest as a Criterion in Vocational Guidance. *Educational Review*, 1916, 52, 349-356.
288. Kitson, H. D. Psychological Tests and Vocational Guidance. *School Review*, 1916, 24, 207-214.
289. Kitson, H. D. The Psychology of Vocational Adjustment. Philadelphia, Lippincott, 1925, pp. 273.
290. Kitson, H. D. Scientific Method in Job Analysis. *Journal of Political Economy*, 1921, 29, 508-514.
291. Kitson, H. D. A Shift of Emphasis Needed in Personnel Research. *Journal of Applied Psychology*, 1922, 6, 141-148.
292. Kitson, H. D., and Donham, C. L. The Personality of Workers in the Metal Trades. A Statistical Study. *Journal of Personnel Research*, 1924, 2, 460-466.
293. Klockenberg, E. A. Beiträge zur Psychotechnik der Schreibmaschine und ihre Beidienung. *Industrielle Psychotechnik*, 1924, 1, 209-246.
294. Klutke, O. Beiträge zur psychotechnischen Eignungsprüfung für den Fernsprehdienst. *Praktische Psychologie*, 1922, 3, 93-110.

295. Klutke, O. Psychotechnische Eignungsprüfung für Funker. *Praktische Psychologie*, 1923, 4, 289-294.
296. Knight, F. B. The Effect of the Acquaintance Factor upon Personal Judgments. *Journal of Educational Psychology*, 1923, 14, 129-142.
297. Knight, F. B., and Fransen, R. H. Pitfalls in Rating Schemes. *Journal of Educational Psychology*, 1922, 13, 204-213.
298. Kohs, S. C. An Ethical Discrimination Test. *Journal of Delinquency*, 1922, 7, 1-15.
299. Kohs, S. C. A Note on Measurement by Relative Position. *Psychological Bulletin*, 1918, 15, 57-60.
300. Kohs, S. C., and Irle, K. Propheying Army Promotion. *Journal of Applied Psychology*, 1920, 4, 73-87.
301. Kornhauser, A. W. A Plan for Apprentice Training. *Journal of Personnel Research*, 1922, 1, 215-230.
302. Kornhauser, A. W. Psychological Test for Non-Office Occupations. *University Journal of Business*, 1924, 2, 175-199.
303. Kornhauser, A. W. Psychological Tests for Office Occupations. *The University Journal of Business (Chicago)*, 1923, 2, 22-51.
304. Kornhauser, A. W. The Psychology of Vocational Selection. *Psychological Bulletin*, 1922, 19, 192-229.
305. Kornhauser, A. W. A Reinterpretation of the Statistical Method of Army Trade Tests. *Journal of Applied Psychology*, 1923, 7, 339-348.
306. Kornhauser, A. W. Scientific Method in Constructing Psychological Tests for Business. *Journal of Political Economy*, 1923, 31, 401-432.
307. Kornhauser, A. W. Some Business Applications of a Mental Alertness Test. *Journal of Personnel Research*, 1922, 1, 103-121.
308. Kornhauser, A. W. A Statistical Study of a Group of Specialized Office Workers. *Journal of Personnel Research*, 1923, 2, 103-123.
309. Kornhauser, A. W., and Jackson, A. W. A Note on the Extent to Which Systems of Character Analysis are Used in the Business World. *Journal of Applied Psychology*, 1922, 6, 302.
310. Kornhauser, A. W., and Kingsbury, F. A. Psychological Tests in Business. Chicago, University Press, 1924, pp. 194.
311. Kornhauser, A. W., and Ruml, B. Recent Developments in Trade Test Theory. *Psychological Bulletin*, 1920, 17, 79-80.
312. Kraus, D. Eignungsprüfungen bei der Einführung von weibliche Ersatzkräften in das Stuttgarter Buchdruckgewerbe. *Zeitschrift für Angewandte Psychologie*, 1918, 13, 121-139.
313. Laemmel, R. Intellegenzprüfung und Psychologische Berufsberatung. Munich, Oldenberg, 1923, pp. 193.
314. Lahy, J. M. Les conditions psychophysiologiques de l'aptitude au travail dactylographique. *Journal de physiologie et pathologie générale*, 1913, 15, 826-835.

315. Lahy, J. M. Une exemple de sélection professionnelle. Les Dactylographes. *Mon. Bureau*, 1920, 3, 159-161.
316. Laird, D. A. The Elements of Personality in Nursing. *American Journal of Nursing*, 1921, 21, 697-702; 775-778.
317. Laird, D. A. The Employee's Past. *Industrial Management*, 1924, 68, 152-160.
318. Laird, D. A. Finding Signs of Success in Employees. *Industrial Management*, 1925, 69, 372-378.
319. Laird, D. A. Personal Data and Vocational Fitness. *Industrial Management*, 1925, 69, 163-171.
320. Laird, D. A. Pitfalls of the Employment Interview. *Industrial Management*, 1924, 68, 58-62.
321. Laird, D. A. Psychology Dons Overalls. *Industrial Management*, 1922, 66, 204-211.
322. Laird, D. A. The Psychology of Selecting Men. New York, McGraw, 1925, pp. 274.
323. Laird, D. A. Recording the Worker's Personality. *Industrial Management*, 1924, 68, 307-314.
324. Laird, D. A. What is the Letter of Application Worth? *Industrial Management*, 1923, 66, 336-344.
325. Laird, D. A., and Remmers, H. A. Study of Estimates of Intelligence from Photographs. *Journal of Experimental Psychology*, 1924, 7, 429-446.
326. Lamb, J. P. Intelligence Tests in Industry. *Industrial Management*, 1919, 58, 21-23.
327. Landis, C. The Justification of Judgments. *Journal of Personnel Research*, 1925, 4, 7-19.
328. Landis, M. H., and Burtt, H. E. A Study of Conversations. *Journal of Comparative Psychology*, 1924, 4, 81-89.
329. Leeming, J. Necessary Physical and Mental Requirements of Platform Employees. *Electric Railway Journal*, 1922, 59, 404-406.
330. Leibenberg, R. Methode und Technik der Berufsberatung. Leipzig, Quelle und Meyer, 1925, pp. 154.
331. Lewis, B. C. Choosing a State Official by Mental Tests. *American Review of Reviews*, 1922, 66, 171-176.
332. Lichtenberger, J. P. Social Significance of Mental Levels. *Publications of American Sociological Society*, 1921, 15, 102-124.
333. Lichtner, W. O. Time Study and Job Analysis. New York, Ronald, 1921, pp. 397.
334. Liebmann, H. Eignungsprüfung von Vorkalkulatoren bei der Badischen Anilin-Sodafabrik. *Industrielle Psychotechnik*, 1924, 1, 137-140.
335. Link, H. The Applications of Psychology to Industry. *Psychological Bulletin*, 1920, 17, 335-346.

336. Link, H. C. *Education and Industry*. New York, Macmillan, 1923, pp. 265.
337. Link, H. C. *Employment Psychology*. New York, Macmillan, 1919, pp. 440.
338. Link, H. C. An Experiment in Employment Psychology. *Psychological Review*, 1918, 25, 116-127.
339. Link, H.-C. Personal Opinion, Records and Rating Scales. *Industrial Management*, 1923, 66, 78-80.
340. Link, H. C. Psychological Tests in Industry. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 32-44.
341. Lipmann, O. Bibliographie zur psychologischen Berufsberatung, Berufseignungsforschung und Berufskunde. *Schriften zur Psychologie der Berufseignung und des Wirtschaftslebens*, 1922, 20, 1-49.
342. Lipmann, O. Die Berufseignung der Schriftsetzer. *Zeitschrift für Angewandte Psychologie*, 1918, 13, 105-120.
343. Lipmann, O. Die psychische Eignung der Funkel-telegraphers. *Zeitschrift für Angewandte Psychologie*, 1919, 15, 301-340.
344. Lipmann, O. The Principles of Vocational Guidance. *British Journal of Psychology*, 1924, 14, 321-325.
345. Lipmann, O., and Stolzenberg, O. Methoden zur Auslese hochwertiger Facharbeiter der Metallindustrie. *Zeitschrift für Angewandte Psychologie*, 1920, 16, 173-249.
346. Lowell, A. L. College Studies and the Professional School. *Educational Review*, 1911, 42, 217-233.
347. McCants, M. Tests Used in Selecting Employees. *Electric Railway Journal*, 1922, 60, 679-681.
348. McClain, J. An Evaluation of the Meeting Objections Tests. Unpublished Thesis. Carnegie Institute of Technology, 1923.
349. McComas, H. C. Some Tests for Efficiency of Telephone Operators. *Journal of Philosophy, Psychology and Scientific Methods*, 1914, 11, 293-294.
350. McDaniel, H. B. Minimum Specifications for Highway Engineering Positions. *Bulletin of the National Research Council*, 1924, 8, part 3, no. 45, pp. 1-105.
351. McGrath, M. C. A Study of the Moral Development of Children. *Psychological Monographs*, 1923, 32, no. 2, pp. 161.
352. McKail, D. Discussion on Industrial Fatigue and Vocational Selection on a Basis of Physical Inquiry. *British Medical Journal*, 1922, whole no. 3217, pp. 351-355.
353. Maday, S. V. Zur Eignungsprüfung von höheren Kanzleirangestellten (Konzipienten). *Zeitschrift für Angewandte Psychologie*, 1919, 14, 306-307.
354. Mandel, A. Getting and Keeping Good Policemen. *National Municipal Review*, 1923, 12, 299-302.

355. Mann, A. L. Placement of Operators Through Tests. *Personnel Administration*, 1922, 10, 10-17.
356. Mann, C. R. A Study of Engineering Education. Carnegie Foundation for the Advancement of Teaching, Bulletin no. 11, 1918, pp. 139.
357. Manson, G. E. Bibliography on Psychological Tests and Other Objective Measures in Industrial Personnel. *Journal of Personnel Research*, 1925, 4, 301-325.
358. Manson, G. E. Group Differences in Intelligence Tests. *Journal of Applied Psychology*, 1925, 9, 156-175.
359. Manson, G. E. The Personal History Record. Typed copies on file in Carnegie Library, Pittsburgh.
360. Manson, G. E. What Can the Application Blank Tell? *Journal of Personnel Research*, 1925, 4, 73-99.
361. Marcus, L. Vocational Selection for Specialized Tasks. *Journal of Applied Psychology*, 1920, 4, 186-201.
362. Martens, H. A. Psychologie und Verkehrswesen. *Zeitschrift für Angewandte Psychologie*, 1919, 15, 374-385.
363. Martin, E. M. An Aptitude Test for Policemen. *Journal of Criminal Law and Criminology*, 1923, 14, 376-404.
364. Martin, E. M. New Methods of Selecting Policemen. *National Municipal Review*, 1923, 12, 671-681.
365. Masterson, J. L. The Art of Interviewing. *Industrial Management*, 1920, 60, 247-248.
366. Maxwell, Sadie. New Tests for Clerical Workers. *Office Economist*, 1924, 6, 5-6.
367. May, M. A. The Present Status of the Will Temperament Tests. *Journal of Applied Psychology*, 1925, 9, 29-52.
368. Mayo, E. The Basis of Industrial Psychology. *Bulletin of the Taylor Society*, 1924, 9, 249-259.
369. Mayo, E. The Irrational Factor in Human Behavior. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 117-130.
370. Meine, F. Job Analysis for Employment Purposes. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 22-31.
371. Meine, F. Job Specifications. Federal Board for Vocational Education, Bulletin 45, 1919, pp. 63.
372. Meine, F., Staud, R. W., and Stearns, W. D. Report of Committee on Descriptions of Occupations of the Industrial Relations Association, 1920. (Multigraphed.)
373. Merrill, M. A. The Relation of Mental Age to Industrial Efficiency of a Group of Mental Defectives. *Journal of Delinquency*, 1925, 9, 83-104.

374. Metcalf, H. C. (editor). *Linking Science and Industry*. Baltimore, Williams and Wilkins, 1925, pp. 206.
375. Metcalf, H. C. Report of Committee on Vocational Guidance. National Association of Corporation Schools, New York, 1916, 94-111.
376. Metcalf, H. C., and Tead, O. Applying Selective Tests to Industry — Essential Constructive Qualities. *Personnel*, 1919, 1, no. 6, p. 2.
377. Miles, G. H. The Berlin Conference on Applied Psychology. *Journal of the National Institute of Industrial Psychology*, 1923, 1, 190-192.
378. Miles, G. H. Economy and Safety in Transport. *Journal of the National Institute of Industrial Psychology*, 1925, 2, 192-197.
379. Miles, G. H., and Angeles, A. An Investigation in a New Industry. *Journal of the National Institute of Industrial Psychology*, 1923, 1, 275-276.
380. Miles, G. H., Eyre, A. H. B., and Bennet, H. P. An Investigation in a Cabinet Factory. *Journal of the National Institute of Industrial Psychology*, 1923, 1, 223-227.
381. Mills, J. Engineering Aptitudes. An Interviewer's Method of Determining Basic Aptitudes of Engineering Graduates. *Journal of Personnel Research*, 1924, 3, 197-206.
382. Miner, J. B. An Aid to the Analysis of Vocational Interests. *Journal of Educational Research*, 1922, 5, 311-323.
383. Miner, J. B. The Evaluation of a Method for Finely Graded Estimates of Abilities. *Journal of Applied Psychology*, 1917, 1, 123-133.
384. Miner, J. B. Standardizing Test for Vocational Guidance. *School and Society*, 1921, 13, 629-633.
385. Mitchell, D., and Ruger, G. J. *Psychological Tests. Revised and Classified Bibliography*. New York, Bureau of Educational Measurements, 1918, pp. 116.
386. Moede, W. Die Eignungsprüfung im Dienste der Betriebsrationalisierung. *Industrielle Psychotechnik*, 1924, 1, 1-16.
387. Moede, W. Ergebnisse der Industriellen Psychotechnik. *Praktische Psychologie*, 1921, 2, 289-329.
388. Moede, W. Die Experimentalpsychologie im Dienste des Wirtschaftslebens. Berlin, Springer, 1919, pp. 111.
389. Moede, W. Frage und Beobachtungsbogen in der praktischen Psychologie. *Praktische Psychologie*, 1923, 4, 131-164.
390. Moede, W. Die Psycho Technische Arbeits Studie. Richtlinien für die Praxis. *Praktische Psychologie*, 1920, 1, 135-146.
391. Moede, W. Die psychotechnische Eignungsprüfung des Industriellen Lehrlings. Aufmerksamkeit und Auftrags erledigung. *Praktische Psychologie*, 1919, 1, 65-81.
392. Moede, W. Psychotechnische Eignungsprüfung in der Industrie. *Praktische Psychologie*, 1920, 1, 339-350, 365-371.

393. Moll, A., and Piorkowski, C. Psychotechnische Prüfung von Kriminalbeamten. *Praktische Psychologie*, 1920, 1, 309-319.
394. Moore, B. V. Personnel Selection of Graduate Engineers. *Psychological Monographs*, 1921, 30, whole no. 138, pp. 85.
395. Moore, B. V. Preventing Misfits by Applied Psychology. *Industrial Management*, 1923, 66, 59-62.
396. Moore, B. V. A Tested Method of Using Tests for Vocational Guidance. *School and Society*, 1923, 18, 761-764.
397. Moore, B. V. Vocational Placement of Graduate Engineers. *Service Bulletin, Bureau of Personnel Research*, Carnegie Institute of Technology, 1922, 5, no. 2, pp. 16.
398. Moore, H. T., and Gilliland, A. R. The Measurement of Aggressiveness. *Journal of Applied Psychology*, 1921, 5, 97-118.
399. Moss, F. A. Standardized Tests for Automobile Drivers. *Public Personnel Studies*, 1925, 3, 147-165.
400. Moss, F. A. Suggested Tests for Hospital Attendant. *Public Personnel Studies*, 1924, 2, 167-178.
401. Moss, F. A. Suggested Tests for Painter. *Public Personnel Studies*, 1925, 3, 131-136.
402. Moss, F. A., and Allen, H. H. The Personal Equation in Automobile Driving. *Journal of Society of Automotive Engineers*, 1925, 16, 415-420.
403. Moss, F. A., and Telford, F. Suggested Tests for Food Inspector. *Public Personnel Studies*, 1924, 2, 254-263.
404. Moss, F. A., and Telford, F. Suggested Tests for Senior Clerk. *Public Personnel Studies*, 1924, 2, 195-213.
405. Münsterberg, H. Business Psychology. Chicago, La Salle Extension University, 1915, pp. 296.
406. Münsterberg, H. Grundzüge der Psychotechnik. Leipzig, Barth, 1920, pp. 769.
407. Münsterberg, H. Psychologie und Wirtschaftsleben. Leipzig, Barth, 1912.
408. Münsterberg, H. Psychology and Industrial Efficiency. Boston, Houghton Mifflin Company, 1913, pp. 316.
409. Münsterberg, H. Vocation and Learning. St. Louis People's University, 1922, pp. 289.
410. Muscio, B. Lectures on Industrial Administration. London, Routledge, 1920, pp. 300.
411. Muscio, B. Lectures on Industrial Psychology. London, Routledge, 1920, pp. 300.
412. Muscio, B. The Psycho-physiological Capacities Required by the Hand Compositor. Industrial Fatigue Research Board, Report 16, sect. A.
413. Muscio, B. Vocational Guidance. A Review of the Literature. Industrial Fatigue Research Board, Report 12.

414. Muscio, B., and Brooke, R. St. C. An Investigation into the Packing of Chocolates (II). *Journal of the National Institute of Industrial Psychology*, 1922, 1, 68-75.
415. Muscio, B., and Sowton, S. C. Vocational Tests and Typewriting. *British Journal of Psychology*, 1923, 13, 344-369.
416. Myers, C. S. *Mind and Work*. New York, Putnam, 1921, pp. 175.
417. Myers, C. S. Psychology and Industry. *British Journal of Psychology*, 1920, 10, 177-182.
418. Myers, C. S. The Use of Psychological Tests. *Proceedings of Association for the Advancement of Education in Industry and Commerce*, 1921, 4, 28-37.
419. National Academy of Sciences, Memoirs. *Psychological Examining in the United States Army*, 1921, 15, pp. 890.
420. National Association of Corporation Schools. Psychological Tests and Results Obtained Therefrom. *Proceedings*, 1919, 7, 342-366.
421. National Association of Corporation Schools. Report of Committee on Application of Psychological Tests and Rating Scales in Industry. *Proceedings*, 1920, 8, 107-153; 1921, 9, 507-556.
422. National Association of Corporation Schools. Report of Committee on Job Analysis. *Proceedings*, 1920, 8, 154-199; 1921, 9, 201-241.
423. National Institute of Industrial Psychology. The Institute's Evidence before the Committee on Industry and Trade. *Journal of the National Institute of Industrial Psychology*, 1925, 2, 283-290.
424. Nicholson, F. W. Success in College and in After Life. *School and Society*, 1915, 2, 229-232.
425. Norris, H. H. Picking Men for Jobs in the Transportation Department. *Electric Railway Journal*, 1921, 58, 607-608.
426. Norsworthy, N. The Validity of Judgments of Character. *Essays in Honor of William James*, 1910, pp. 542-552.
427. O'Connor, J. General Electric Develops Practical Tests. *Babson's Reports*, Sept., 1924, pp. 6-7.
428. Omwake, K. The Value of Photographs and Handwriting in Estimating Intelligence. *Public Personnel Studies*, 1925, 3, 2-15.
429. Ordahl, G. Industrial Efficiency of the Moron. *Training School Bulletin*, 1919, 15, 145-152.
430. O'Rourke, L. J. Saving Dollars and Energy by Personnel Research. *Journal of Personnel Research*, 1926, 4, 351-364.
431. Ortmann, O. Weight Discrimination as a Measure of Technical Skill in Piano Playing. *Journal of Comparative Psychology*, 1923, 3, 11-26.
432. Oschrin, E. Vocational Tests for Retail Saleswomen. *Journal of Applied Psychology*, 1918, 2, 148-155.
433. O'Shea, P. F. Fitting Workers into Specialized Jobs. *Factory*, 1920, 24, 1884-1888.



434. Otis, A. S. Mental Tests and Manual Labor; Selection of Mill Workers by Mental Tests. *Scientific American Monthly*, 1921, 4, 271.
435. Otis, A. S. The Selection of Mill Workers by Mental Tests. *Journal of Applied Psychology*, 1920, 4, 339-341.
436. Packard Motor Car Company. Placing the New Employee. *Iron Age*, 1919, 103(2), 1289.
437. Paterson, D. G. The Graphic Rating Scale. *Journal of Personnel Research*, 1922, 1, 361-376.
438. Paterson, D. G. Methods of Rating Human Qualities. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 81-93.
439. Paterson, D. G. The Scott Company's File Clerk's Test. *Journal of Personnel Research*, 1923, 1, 547-561.
440. Paterson, D. G. The Vocational Testing Movement. *Journal of Personnel Research*, 1922, 1, 295-305.
441. Paterson, D. G., and Ludgate, K. Blonde and Brunette Traits; a Quantitative Study. *Journal of Personnel Research*, 1922, 1, 122-127.
442. Paterson, D. G., and Ruml, B. The Extension of Rating Scale Theory and Technique. *Psychological Bulletin*, 1920, 17, 80-81.
443. Patten, E. F. An Experiment in Testing Engine Lathe Aptitude. *Journal of Applied Psychology*, 1923, 7, 16-29.
444. Payne, A. F. Measuring of Trade Knowledge and Trade Skill. *Industrial Arts Magazine*, 1920, 9, 379-381.
445. Payne, A. F. The Scientific Selection of Men. *Scientific Monthly*, 1920, 11, 544-547.
446. Piéron, H. Le rôle de la psychologie dans l'orientation professionnelle. *Notes et Mémoires l'institut Lannelongue*, 1923, 4, 42-76.
447. Pintner, R. Intelligence as Estimated from Photographs. *Psychological Review*, 1918, 25, 286-296.
448. Pintner, R. Intelligence Testing. New York, Holt, 1923, pp. 406.
449. Pintner, R., and Paterson, D. G. A Scale of Performance Tests. New York, Appleton, 1917, pp. 218.
450. Piorkowski, C. Beiträge zur Psychologischen Methodologie der wirtschaftlichen Berufseignung. *Zeitschrift für Angewandte Psychologie*, 1916, 11, -84.
451. Piorkowski, C. Über eine Angestelltenprüfung bei der Auerlicht-Gesellschaft. *Praktische Psychologie*, 1919, 1, 33-40.
452. Poffenberger, A. T. A Critical Examination of the Usual Employment Methods. *Annals of the American Academy of Political and Social Science*, 1923, 110, No. 199, 13-21.
453. Poffenberger, A. T. Psychological Examinations; New and Practical Way of Measuring Vocational Fitness. *Scientific American Monthly*, 1921, 3, 205-211.

454. Poffenberger, A. T. Psychological Tests for Vocational Guidance. *Nature*, 1921, 108, 321-323.
455. Poffenberger, A. T. Psychology as a Business. *Scientific American*, 1922, 127, 112-113, 142.
456. Poffenberger, A. T. School Achievement and Success in Life. *Journal of Applied Psychology*, 1925, 9, 22-28.
457. Poffenberger, A. T. The Selection of a Successful Secretary. *Journal of Applied Psychology*, 1922, 6, 155-160.
458. Poffenberger, A. T., and Vartanian, V. H. The Letter of Application in Vocational Selection. *Journal of Applied Psychology*, 1922, 6, 74-80.
459. Pope, D. V. The Interpretation of the Human Face from Photographs. *Bulletin of Randolph Macon Woman's College*, 1922, 8, 3-17.
460. Porter, H. F. Letting the Worker Pick His Job. *Factory*, 1921, 26, 1306-1307.
461. Prankel. Ergebnisse einer psychologische Eignungsprüfung bei der Eisenbahnlokomotivwerkstätte Gleiwitz. *Zeitschrift für Angewandte Psychologie*, 1920, 16, 393-344.
462. Pressey, S. L., and L. C. Introduction to the Use of Standard Tests. Yonkers, World Book Co., 1922, pp. 263.
463. Proctor, W. M., and Ward, H. Relation of General Intelligence to the Persistence of Educational and Vocational Plane of High School Pupils. *Journal of Educational Research*, 1923, 7, 277-288.
464. Pruette, L. What Mental Tests Can Do for Industry. *Industrial Management*, 1925, 70, 270-271.
465. Pruette, L., and Fryer, D. Group Problems of Executives. *Journal of Personnel Research*, 1924, 3, pp. 39-45.
466. Public Personnel Studies. Suggested Tests for Fire Fighter. *Public Personnel Studies*, 1924, 2, 226-234.
467. Quigley, H. Industrial Psychology in Germany. *Psyche*, 1923, 3, 229-234.
468. Ream, M. J. Ability to Sell. Baltimore, Williams and Wilkins, 1924, pp. 64.
469. Ream, M. J. Graphic Rating Scale for Insurance Salesmen. Unpublished thesis. Carnegie Institute of Technology, 1921, pp. 85-86.
470. Ream, M. J. A Group Will Temperament Test. *Journal of Educational Psychology*, 1922, 13, 7-16.
471. Ream, M. J. The Prediction of Successful Salesmanship. (Type-written.) On file in Carnegie Library, Pittsburgh, 1921.
472. Ream, M. J. A Social Relations Test. *Journal of Applied Psychology*, 1922, 6, 69-73.
473. Ream, M. J. A Statistical Method for Incomplete Order of Merit Ratings. *Journal of Applied Psychology*, 1921, 5, 261-266.

474. Reasoner, M. H. Doing Half the Job Before You Start. *Factory*, 1920, 25, 552-554.
475. Reilly, P. J. Rating Scale in Industry. *Personnel*, 1919, 1, no. 7, pp. 1, 3, and 6.
476. Reilly, P. J. The Selection and Placement of Employees. U.S. Federal Board for Vocational Education, Bulletin 49, 1920, pp. 83.
477. Richmond, H. A. Determining Factors in Selection of Salesmen. *Administration*, 1923, 5, 49-58.
478. Richmond, H. A. Research in Sales Personnel. Bureau of Personnel Research, Carnegie Institute of Technology, Annual Report, 1921.
479. Richmond, H. A. Selecting Employees by Tests. *Industrial Management*, 1920, 60, 150-152.
480. Ringo, F. J. What Buyers and Assistant Buyers Need to Know. *Journal of Personnel Research*, 1924, 3, 235-243.
481. Robert, A. Examen Psychotechnique des Pontonniers. Rep. Deuxième Conférence Internationale de Psychotechnique Appliquée à l'Orientation Professionnelle et à l'Organisation Scientifique du Travail. Barcelona, 1922.
482. Roberts, E. B. Selecting and Placing Technical Graduates in Westinghouse Organization. *Management and Administration*, 1923, 6, 207-211.
483. Robinson, B. W. Capacity and Aptitude for Stenography and Typing. (Unpublished Thesis.) University of Iowa, 1921.
484. Robinson, E. S. The Analysis of Trade Ability. *Journal of Applied Psychology*, 1919, 3, 352-357.
485. Roderick, A. H. A Practical System for Rating Shop Employees. *Factory*, 1924, 32, 32-33.
486. Roemer, G. A. Prüfung von Kriminalanwärtern. *Zeitschrift für Angewandte Psychologie*, 1921, 18, 107-111.
487. Roemer, G. A. Psychographische Tiefenanalyse eines Grossindustriellen und seines Stabes. *Praktische Psychologie*, 1922, 4, 16-28.
488. Rogers, H. W. Psychological Tests for Stenographers and Typewriters. *Journal of Applied Psychology*, 1917, 1, 268-274.
489. Rogers, H. W. Some Empirical Tests in Vocational Selection. *Archives of Psychology*, 1922, no. 49, pp. 47.
490. Rosenow, C. The Analysis of Mental Functions. *Psychological Monographs*, 1917, 25, whole no. 106, pp. 43.
491. Rossi, W. H., and D. I. Personnel Administration: A Bibliography. Baltimore, Williams and Wilkins, 1925, pp. 365.
492. Rossy, C. S. Mental Examinations for Employees. *Industrial Management*, 1917, 54, 359-365.
493. Rowe, D. E. Job Analysis as an Aid to Cost Reduction. *Industrial Management*, 1922, 63, 341-346.
494. Rowntree, J. S. The Scope of Vocational Selection in Industry.

- Journal of the National Institute of Industrial Psychology*, 1923, 1, 240-245.
495. Ruch, G. M., and Del Manzo, M. C. The Downey Will Temperament Group Test — A Further Analysis of Its Reliability and Validity. *Journal of Applied Psychology*, 1923, 7, 65-76.
496. Ruffer, W. Auswertungserfahrungen der Psychotechnischen Prüfstelle der Osram G.m.b. Kommanditgesellschaft, Fabrik S. *Praktische Psychologie*, 1923, 4, 225-236.
497. Rugg, H. O. Is the Rating of Human Character Practicable? *Journal of Educational Psychology*, 1921, 12, 425-438, 485-501; 1922, 13, 30-42, 81-93.
498. Rugg, H. O. *Statistical Methods Applied to Education*. Boston, Houghton Mifflin Company, 1917, pp. 410.
499. Ruggles, A. M. A Diagnostic Test of Aptitude for Clerical Office Work. New York, Teachers College, Columbia University Contributions to Education, no. 148, 1924, pp. 85.
500. Ruml, B. Extension of Selective Tests to Industry. *Annals of the American Academy of Political and Social Science*, 1919, 81, 38-46.
501. Rupp, H. Eignungsprüfungen für Telephonistinnen. *Zeitschrift für Angewandte Psychologie*, 1921, 19, 63-76.
502. Rupp, H. Wie Entwickelt Man Psychologische Eignungsprüfungen? *Betrieb*, 1921, 4, 96-105.
503. Russell Sage Foundation Library. *Psychological Tests in Industry*, Bulletin 49, 1921, pp. 4.
504. Sachs, H. Studien zur Eignungsprüfung der Strassenbahnführer. *Zeitschrift für Angewandte Psychologie*, 1921, 17, 199-225.
505. Schachwitz, A. Über psychologische Eignungsprüfungen für Verkehrsberufe. Eine Begutachtung ihres theoretischen und praktischen Wertes erläutert durch eine Untersuchung von Strassenbahnführern. Berlin, Julius Springer, 1920, pp. 187.
506. Schlesinger, H. Praktische Ergebnisse aus der industriellen Psychotechnik. *Zeitschrift des Vereines deutscher Ingenieure*, 1920, 64, 417-424.
507. Schlichling, P. Die Auswahl von Mechanikerlehrlingen bei der Firma Wilh. Morrell in Leipzig. *Praktische Psychologie*, 1920, 2, 21-26.
508. Schneickert, H. Die Psychotechnische Methode bei der Auswahl der Tüchtigsten, insbesondere im Kriminaldienste. *Deutsche Strafrechts-Zeitung*, 1920, 7, 221-225.
509. Schneider, H. Eignungsprüfung und Erfolgskontrolle in einem Grossbetriebe der Elektroindustrie. *Industrielle Psychotechnik*, 1925, 2, 108-118.
510. Schneider, H. Selecting Men for Particular Work. *National Association of Corporation Schools*, Bulletin 3, 1916, pp. 23-28.

511. Schneider, H. Selecting Young Men for Particular Jobs. *American Machinist*, 1913, 38, 597-600.
512. Schoen, M. The Nature and Measurement of Musical Talent. *Carnegie Technical Journal*, 1922, 3, 7.
513. Schreiber, I. Das Prüflaboratorium für Berufseignung bei der Eisenbahn-Generaldirektion Dresden. *Praktische Psychologie*, 1921, 2, 232-239.
514. Schroder, F. Berufsanalyse der Buchdrucksparten. *Zeitschrift für Angewandte Psychologie*, 1924, 23, 188-196.
515. Schulte, R. W. Die Berufseignung des Damenfriseurs. *Zeitschrift für Angewandte Psychologie*, 1921, 19, 100-155.
516. Schulte, R. W. Eignungsprüfungen im Friseurgewerbe. *Praktische Psychologie*, 1920, 1, 371-378.
517. Schwarze, B. Beiträge zur Frage der Personalauswahl in Grossbetrieben mit Besonderer Berücksichtigung der Deutschen Reichsbahn. *Industrielle Psychotechnik*, 1924, 1, 17-21.
518. Scott, W. D. A Fourth Method of Checking Results in Vocational Selection. *Journal of Applied Psychology*, 1917, 1, 61-66.
519. Scott, W. D. Measurements of Trade Skill and Intelligence. *Proceedings of First Annual Convention of National Association of Employment Managers*, Cleveland, 1919, 63-67.
520. Scott, W. D. Personnel Administration as an Aid to Industrial Stability. *Annals of the American Academy of Political and Social Science*, 1920, 90, 138-142.
521. Scott, W. D. The Scientific Selection of Salesmen. *Advertising and Selling*, 1915, 25, 5-6, 94-96.
522. Scott, W. D. Selection of Employees by Means of Quantitative Determinations. *Annals of the American Academy of Political and Social Science*, 1916, 65, 182-193.
523. Scott, W. D. Vocational Selection at Carnegie Institute of Technology. U.S. Bureau of Labor Statistics, Bulletin 227, 1917, 114-119.
524. Scott, W. D., Bingham, W. V., and Whipple, G. M. Scientific Selection of Salesmen. *Salesmanship*, 1916, 4, 106-108.
525. Scott, W. D., and Clothier, R. C. Personnel Management. Chicago, Shaw, 1923, pp. 643.
526. Scott, W. D., and Hayes, M. H. S. Science and Common Sense in Working with Men. New York, Ronald, 1921, pp. 154.
527. Seashore, C. E. The Measurement of a Singer. *Science*, 1912, 35, 201-212.
528. Seashore, C. E. The Psychology of Musical Talent. Boston, Silver, Burdett, 1919, pp. 288.
529. Seiffert, G. Die Psychotechnische Prüfung des Gleichgewichtsinnes bei Fliegern. *Praktische Psychologie*, 1919, 1, 81-87.

530. Shaeffer, C. M. Efficiency Tests of Pennsylvania Railroad Personnel. *Journal of Personnel Research*, 1924, 3, 244-251.
531. Shefferman, N. W. Employment Methods. New York, Ronald, 1920, pp. 573.
532. Shellow, S. M. Research in Selection of Motormen in Milwaukee. *Journal of Personnel Research*, 1925, 4, 222-237.
533. Shen, E. The Influence of Friendship Upon Personal Ratings. *Journal of Applied Psychology*, 1925, 9, 66-68.
534. Shen, E. The Validity of Self-Estimate. *Journal of Educational Psychology*, 1925, 16, 104-107.
535. Short, O. C. Suggested Tests for Road Inspector. *Public Personnel Studies*, 1925, 3, 252-265.
536. Short, O. C. Suggested Tests for Shift Engineman. *Public Personnel Studies*, 1925, 3, 320-330.
537. Shuttleworth, F. K. Method of Measuring Character Traits. *School and Society*, 1924, 19, 679-682.
538. Simons, A. M. Personnel Relations in Industry. New York, Ronald, 1921, pp. 341.
539. Slawson, J. The Reliability of Judgments of Personal Traits. 1920. (Typewritten.) On file in Columbia University Library.
540. Smith, M. The Laundry Trade; Experiments with a Psychological Test. *Journal of Industrial Hygiene*, 1924, 6, 14-19.
541. Snow, A. J. An Experiment in the Validity of Judging Human Ability. *Journal of Applied Psychology*, 1924, 8, 339-346.
542. Snow, A. J. Labor Turnover and Mental Alertness Test Scores. *Journal of Applied Psychology*, 1923, 7, 285-290.
543. Snow, A. J. Psychology in Business Relations. Chicago, Shaw, 1925, pp. 562.
544. Snow, A. J. Scientific Selection of Taxicab Drivers. *National Safety News*, 1924, 10, 2-4.
545. Snow, A. J. Tests for Chauffeurs. *Industrial Psychology*, 1926, 1, 30-45.
546. Snow, A. J. Tests for Transportation Pilots. 1925. (Mimeographed.)
547. Spaeth, R. A. Prevention of Fatigue in Industry; Psychological Tests and the Reduction of Necessary Fatigue. *Industrial Management*, 1920, 59, 311-313.
548. Spielman, W. The Vocational Selection of Weavers. *Journal of the National Institute of Industrial Psychology*, 1925, 1, 256-261.
549. Spielman, W. Vocational Tests for Dressmakers' Apprentices. *Journal of the National Institute for Industrial Psychology*, 1923, 1, 277-282.
550. Spielman, W. Vocational Tests for Selecting Packers and Pipers. *Journal of the National Institute of Industrial Psychology*, 1925, 2, 365-372.

551. Stamm, L. E. Medical Aspects of Aviation. 1919. The Royal Aeronautical Society.
552. Starch, D. The Use and Limitations of Psychological Tests. *Harvard Business Review*, 1922, 1, 71-80.
553. Stearns, A. W. The Classification of Industrial Applicants. *American Journal of Insanity*, 1920, 76, 409-417.
554. Stearns, A. W. Note on the Relation between Group Test Score and Capacity. *Journal of Abnormal and Social Psychology*, 1924, 18, 369-374.
555. Stearns, W. D. Placing the Right Man in the Right Job. *Machinery*, 1919, 26, 30-37, 136-139.
556. Steuquist, J. L. Measurements of Mechanical Ability. Teachers' College Contributions to Education 130, Columbia University, 1923, pp. 101.
557. Stern, E. Eignungsprüfung an Schalmesser. *Zeitschrift für Angewandte Psychologie*, 1920, 16, 335-340.
558. Stern, E. Über eine Experimentell-Psychologische Eignungsprüfung für Flugzeugführer. *Zeitschrift für Angewandte Psychologie*, 1919, 15, 236-253.
559. Stern, W. Über eine psychologische Eignungsprüfung für Strassenbahnführerinnen. *Zeitschrift für Angewandte Psychologie*, 1918, 13, 91-104.
560. Stern, W., and Roloff, H. P. Psychologische Auslese der Lehrlinge für deutsche Eisenbahnwerkstätten. *Zeitschrift für Angewandte Psychologie*, 1921, 22, 50-61.
561. Stern-Aachen, E. Materiale Arbeitsanalysen aus der Textilindustrie. *Praktische Psychologie*, 1923, 4, 175-185.
562. Stott, M. Vocational Tests. *Women's Employment*, 1925, 25, 7-8.
563. Stratton, G. M. Psycho-Physical Tests of Aviators. *Scientific Monthly*, 1919, 8, 421-426.
564. Stratton, G. M., McComas, H. C., Coover, J. E., and Bagby, E. Psychological Tests for Selecting Aviators. *Journal of Experimental Psychology*, 1920, 3, 405-423.
565. Streller, J. Die Berufseignung des Mittleren Kaufmannischen Bureau beamten im Buchhandel. Leipzig, Barth, 1921.
566. Streller, J. Erfahrungen mit den Methoden von Lippman-Stotzenberg zur Auslese hochwertiger Facharbeiter der Metall-industrie. *Zeitschrift für Angewandte Psychologie*, 1923, 21, 166-183. (Skilled workers.)
567. Strong, E. K., Jr. Analyzing Industrial Educational Requirements. *Carnegie Technical Journal*, 1921, 2, 17-23.
568. Strong, E. K., Jr. Analyzing Industrial Educational Requirements. *Industrial Education*, 1921, 4, 75-82.

569. Strong, E. K., Jr. Job Analysis of the Manager in Industry. *School and Society*, 1921, 13, 456-462.
570. Strong, E. K., Jr. Psychological Tests and Rating Scales in Relation to Training. New York, National Personnel Association, 1922, pp. 7.
571. Strong, E. K., Jr., and Ubrbrock, R. S. Job Analysis and the Curriculum with Special Reference to the Training of Printing Executives. Baltimore, Warwick and York, 1923, pp. 146.
572. Swan, J. J. Trade Specifications and Occupational Index of Professions and Trades in the Army. Washington, Office of the Adjutant General, War Dept., Document 774, 1918, pp. 239.
573. Swartz, G. O. Selecting Salesmen; Getting the Right Man the First Time. *Journal of Personnel Research*, 1921, 2, 457-459.
574. Symonds, P. M. Notes on Rating. *Journal of Applied Psychology*, 1925, 9, 188-195.
575. Symonds, P. M. On the Loss of Reliability in Ratings Due to Coarseness of the Scale. *Journal of Experimental Psychology*, 1924, 7, 456-461.
576. Tagg, M. Industrial Psychology in Russia. *Journal of the National Institute of Industrial Psychology*, 1925, 2, 359-364.
577. Tagg, M. The "Make-Up" of the Engineering Worker. *Journal of the National Institute of Industrial Psychology*, 1923, 1, 313-324.
578. Tagg, M. The Selection of Engineering Apprentices. *Journal of the National Institute of Industrial Psychology*, 1924, 2, 129-133.
579. Tagg, M. Vocational Tests in the Engineering Trade. *Journal of the National Institute of Industrial Psychology*, 1925, 2, 313-323.
580. Taylor, F. W. The Principles of Scientific Management. New York, Harper Bros., 1911, pp. 144.
581. Tead, O. A Course in Personnel Administration. New York, Columbia University Press, 1923, pp. 257.
582. Tead, O., and Metcalf, H. C. Personnel Administration. New York, McGraw, 1920, pp. 538.
583. Telford, F., and Moss, F. A. Suggested Tests for Janitor. *Public Personnel Studies*, 1925, 3, 26-31.
584. Telford, F., and Moss, F. A. Suggested Tests for Patrolmen. *Public Personnel Studies*, 1924, 2, 112-145.
585. Telford, F., and Moss, F. A. Suggested Tests for Supervising Clerk. *Public Personnel Studies*, 1924, 2, 288-297.
586. Telford, F., and Staff. Suggested Tests for Prison Guard. *Public Personnel Studies*, 1924, 2, 61-101.
587. Telford, F., and Staff. Suggested Tests for Fire Fighters. *Public Personnel Studies*, 1924, 2, 226-234.
588. Terman, L. M. Measurement of Intelligence. Boston, Houghton Mifflin Company, 1916, pp. 362.
589. Terman, L. M. A Trial of Mental and Pedagogical Tests in a Civil



- Service Examination of Policemen and Firemen. *Journal of Applied Psychology*, 1917, 1, 17-29.
590. Terman, L. M. The Use of Intelligence Tests in the Army. *Psychological Bulletin*, 1918, 15, 177-187.
591. Tevis, M. Psychological Tests in Industrial Capacities. *Scientific American Monthly*, 1921, 4, 208-211.
592. Thorndike, E. L. A Constant Error in Psychological Rating. *Journal of Applied Psychology*, 1920, 4, 25-29.
593. Thorndike, E. L. The Correlation between Interests and Abilities in College Courses. *Psychological Review*, 1921, 28, 374-376.
594. Thorndike, E. L. Early Interests: Their Permanence and Relation to Abilities. *School and Society*, 1917, 5, 178-179.
595. Thorndike, E. L. An Important Constant Error in Psychological Rating. *Psychological Bulletin*, 1920, 17, 81-82.
596. Thorndike, E. L. Intelligence and Its Uses. *Harper's Magazine*, 1920, 140, 227-235.
597. Thorndike, E. L. Mental and Social Measurements. New York, Teachers College, 1913, pp. 277.
598. Thorndike, E. L. Scientific Personnel Work in the Army. *Science*, 1919, n.s. 49, 53-61.
599. Thorndike, E. L. The Technique of Combining Incomplete Judgments of the Relative Position of  $N$  Facts Made by  $N$  Judges. *Journal of Philosophy, Psychology and Scientific Methods*, 1916, 13, 197-204.
600. Thurstone, L. L. The Civil Service Tests for Patrolmen in Philadelphia. *Public Personnel Studies*, 1924, 2, 1-5.
601. Thurstone, L. L. A Comparative Study of Clerical Tests. *Public Personnel Studies*, 1923, 1, nos. 2-6.
602. Thurstone, L. L. Employment Tests (office occupations). Yonkers, World Book Co., 1922-1923.
603. Thurstone, L. L. The Fundamentals of Statistics. New York, Macmillan, 1923, pp. 237.
604. Thurstone, L. L. The Intelligence of Policemen. *Journal of Personnel Research*, 1922, 1, 64-74.
605. Thurstone, L. L. Intelligence Tests in the Civil Service. *Journal of Personnel Research*, 1924, 2, 431-441.
606. Thurstone, L. L. Intelligence Tests for Engineering Students. *Engineering Education*, 1923, 13, 263-318.
607. Thurstone, L. L. Mental Tests for Engineering Students. *Proceedings of Society for Promotion of Engineering Education*, 1919, 27, 113-119.
608. Thurstone, L. L. Mental Tests for Prospective Telegraphers. *Journal of Applied Psychology*, 1919, 3, 110-117.
609. Thurstone, L. L. Psychology in the Civil Service. *Annals of the*

- American Academy of Political and Social Science*, 1923, 110, no. 199, 194-199.
610. Thurstone, L. L. Report of Committee No. 22 on Intelligence Tests. *Proceedings of Society for Promotion of Engineering Education*, 1921, 28, 349-356.
  611. Thurstone, L. L. A Scoring Method for Mental Tests. *Psychological Bulletin*, 1919, 16, 235-240.
  612. Thurstone, L. L. A Standardized Test for Office Clerks. *Journal of Applied Psychology*, 1919, 3, 248-251.
  613. Thurstone, L. L. Vocational Guidance Tests. (Five Engineering Tests.) Yonkers, World Book Co., 1922.
  614. Thurstone, L. L. What is Personnel Research? *Journal of Personnel Research*, 1924, 3, 52-56.
  615. Toops, H. A. The Technique of Vocational Guidance as Investigated by the Institute of Educational Research. *National Vocational Guidance Association Bulletin*, 1923, no. 6, 86-93.
  616. Toops, H. A. Trade Tests in Education. *Teachers College Contributions to Education*, no. 115, 1921, pp. 118.
  617. Tramm, K. A. Die Bewährung des psychotechnischen Prüfverfahrens für Strassenbahnführer. *Industrielle Psychotechnik*, 1924, 1, 36-42.
  618. Tramm, K. A. Die rationelle Ausbildung des Fahrpersonals für Strassenbahnen auf psychotechnischer Grundlage. *Praktische Psychologie*, 1919, 1, 18-33.
  619. Troland, L. T. A Technique for the Experimental Study of Telepathy and other Alleged Clairvoyant Processes. 1917, pp. 26. (Published privately.)
  620. Tuttle, W. W. The Determination of Ability for Learning Typewriting. *Journal of Educational Psychology*, 1923, 14, 177-181.
  621. United States Bureau of Labor Statistics. Descriptions of Occupations. *Monthly Labor Review*, 1919, 8, 441-443.
  622. United States Civil Service Commission. Annual Report of the Chief Examiner and the Director of Research, 1923, pp. 64.
  623. United States War Department. Army Mental Tests, Methods, Typical Results, and Practical Applications. Washington, 1918.
  624. United States War Department. Personnel System of the United States Army (2 vols.), Washington, 1919, pp. 713; 342.
  625. United States War Department, Committee on Classification of Personnel in the Army. Trade Test Division. Trade Tests. Washington, 1918. (Not published for general distribution.)
  626. Uhrbrock, R. S. History of Job Analysis. *Administration*, 1922, 3, 164-168.
  627. Uhrbrock, R. S. Job Analysis as a Factor in Cost Reduction. *Industrial Management*, 1922, 63, 100-102.

628. Uhrbrock, R. S. Vocational Psychographs. *Education*, 1921, 41, 510-515.
629. Uyeno, Y. Present Status of Industrial Psychology in Japan. *Psychological Bulletin*, 1924, 21, 225-226.
630. Valentine, T. Zur experimentellen Feststellung von berufswichtigen Willenseigenschaften bei Jugendlichen. *Praktische Psychologie*, 1923, 4, 10-16.
631. Van Deventer, J. H. More Work per Man; Tested and Selected Methods of Managing Men. New York, The Engineering Magazine Company, 1921, pp. 440.
632. Viteles, M. S. The Clinical Viewpoint in Vocational Selection. *Journal of Applied Psychology*, 1925, 9, 131-138.
633. Viteles, M. S. Job Specifications and Diagnostic Tests of Job Competency Designed for the Auditing Division of a Street Railway Company. *Psychological Clinic*, 1922, 14, 83-105.
634. Viteles, M. S. Psychology in Business in England, France, and Germany. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 207-220.
635. Viteles, M. S. Research in Selection of Motormen. *Journal of Personnel Research*, 1925, 4, 100-115, 173-199.
636. Viteles, M. S. Selecting Cashiers and Predicting Length of Service. *Journal of Personnel Research*, 1924, 2, 467-473.
637. Viteles, M. S. Tests in Industry. *Journal of Applied Psychology*, 1921, 5, 57-63.
638. Viteles, M. S. Vocational Guidance and Job Analysis. *Psychological Clinic*, 1924, 15, 157-182.
639. Voelker, P. F. The Functions of Ideals and Attitudes in Social Education. New York, Teachers College, Columbia University, Contributions to Education no. 112, 1921, pp. 126.
640. Vollmer, A. A Practical Method for Selecting Policemen. *Journal of Criminal Law and Criminology*, 1921, 11, 571-582.
641. Waldau, M. Psychotechnische Eignungsprüfungen von anzulernenden Arbeiterinnen der elektrotechnischen Massenherstellung. *Betrieb*, 1921, 4, 110-117.
642. Waldau, M. Psychotechnische Eignungsprüfung von Facharbeiterinnen in der Elektroindustrie. *Praktische Psychologie*, 1921, 2, 160.
643. Waldo, R. J. Cutting Costs by Special Employment Tests. *Administration*, 1921, 2, 739-747.
644. Walton, L. C. A Study of Judgments of Letters of Application. (Master's Essay on file in Columbia University Library.)
645. Washburn, M. F., and Stepanova, V. Absolute Judgments of Character Traits in Self and in Others. *American Journal of Psychology*, 1923, 34, 96-99.
646. Watson, J. B. The Possibilities and Limitations of Psychology in the

- Office. *Proceedings, Third National Conference National Association Office Managers*, 1922, 4-10.
647. Watson, M. Trade Test Principles in Industry. *Industrial Management*, 1920, 59, 247-250.
648. Watts, F. The Construction of Tests for the Discovery of Vocational Fitness. *Journal of Applied Psychology*, 1921, 5, 240-251.
649. Watts, F. An Introduction to the Psychological Problems of Industry. London, Allen and Unwin, 1921, pp. 240.
650. Watts, F. The Outlook for Vocational Psychology. *British Journal of Psychology*, 1921, 11, 194-205.
651. Weakly, F. E. Applied Personnel Procedure. New York, McGraw, 1923, pp. 192.
652. Weakly, F. E. Personnel and Medical Audit; the Selection, Testing, Training and Welfare of Employees. *Annals of the American Academy of Political and Social Science*, 1917, 71, 117-135.
653. Weaver, A. T. Experimental Studies in Vocal Expression; The Prediction of Talent for Vocal Expression in Reading. *Journal of Applied Psychology*, 1924, 8, 159-186.
654. Weaver, G. G. Trade Tests, Their Construction, Use and Possibilities in Industry. *Industrial Arts Magazine*, 1921, 10, 163-166.
655. Weber, C. O. The Psychology of Employment. *Administration*, 1922, 3, 672-676.
656. Weber, J. Die Eignungsprüfung bei Rheinmetall in Dusseldorf. *Psychologische Mitteilungen*, 1921, 2, 145-150.
657. Wells, F. L. Analysis of a Successful Agent. *Life Association News*, 11, 3.
658. Wells, F. L. Briefier Studies from the Psychological Laboratory of McLean Hospital. *Journal of Applied Psychology*, 1919, 3, 172-193.
659. Wembridge, H. A. Experiment and Statistics in the Selection of Employees. *Journal of the American Statistical Association*, 1923, 18, 600-606.
660. Wembridge, H. A. Experiment in Picking Employees. *American Industries*, 1923, 24, 18-20.
661. Werner, H. Prüfung der Fähigkeit der Geschwindigkeitsschätzung und Bremsführung an Triebwagenfuhrern. *Praktische Psychologie*, 1923, 4, 113-117.
662. Whipple, G. M. Manual of Mental and Physical Tests. 2 vols. Baltimore, Warwick and York, 1915, pp. 688.
663. Whipple, G. M. The Use of Mental Tests in Vocational Guidance. *Annals of the American Academy of Political and Social Science*, 1916, 65, 193-204.
664. Weinand, C. M. One Industry's Attitude toward Selection by Mental Levels. *Journal of Educational Psychology*, 1924, 15, 180-182.

665. Willet, G. W. Permanence of Pupil Interests. *School and Society*, 1919, 9, 334-338, 365-368.
666. Willits, J. Systematic Selection of Employees. *Electrical World*, 1923, 81, 601-602.
667. Wilson, D. R. Some Recent Reports on Selection in England; Abstract. *Journal of Industrial Hygiene*, 1922, 4, 54-57.
668. Wood, C., Jr. The Printing Trade in Philadelphia. *Psychological Clinic*, 1924, 15, 182-202.
669. Woodruff, W. H. Forms for Follow-Up of New Employees. *Journal of Personnel Research*, 1924, 3, 289-291.
670. Woodruff, W. H. Standardized Stenographic Tests. *Journal of Personnel Research*, 1923, 1, 562-567.
671. Woodruff, Ruth. A Trade and Its Training. The Dental Laboratory Mechanic. *Journal of Personnel Research*, 1922, 1, 159-170.
672. Woods, E. A. Selecting and Training Agents. Report of the Committee on Scientific Salesmanship of the National Association of Life Underwriters. *Life Insurance Courant*, Oct., 1922, pp. 773-778.
673. Woodworth, R. S. Psychological Experience with the Interview. *Journal of Personnel Research*, 1925, 4, 162-165.
674. Woodworth, R. S., and Wells, F. L. Association Tests. *Psychological Monographs*, 1911, 13, whole no. 57, pp. 85.
675. Woolley, H. T., and Fischer, C. R. Mental and Physical Measurements of Working Children. *Psychological Monographs*, 1914, 18, whole no. 77, pp. 247.
676. Wylie, A. T. Army Experiments with Progress and Proficiency Tests. *Educational Record*, 1924, 5, 254-257.
677. Yerkes, R. M. Mental Tests in Industry. *Transactions of the American Institute of Mining and Metallurgical Engineers*, 1919, 60, 752-766.
678. Yerkes, R. M. Practical Mental Measurement. *Scientific American Monthly*, 1920, 1, 270-273.
679. Yerkes, R. M. Psychological Work of the National Research Council. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 172-178.
680. Yerkes, R. M. Report of the Psychology Committee of the National Research Council. *Psychological Review*, 1919, 26, 83-149.
681. Yerkes, R. M. Testing the Human Mind. *Atlantic Monthly*, 1923, 131, 358-370.
682. Yerkes, R. M. What is Personnel Research? *Journal of Personnel Research*, 1922, 1, 56-63.
683. Yoakum, C. S. Applications of Psychology to Business and Industry. *Journal of Personnel Research*, 1922, 1, 306-318.
684. Yoakum, C. S. Basic Experiments in Vocational Guidance. *Journal of Personnel Research*, 1922, 1, 18-34.

## BIBLIOGRAPHY

563

685. Yoakum, C. S. Can Executives be Picked by Mental Tests? *Forbes*, 1922, 9, 259-260.
686. Yoakum, C. S. Present Limitations of Psychological Tests in Management. American Management Association (Annual Convention Series: no. 21), 1925, pp. 3-14.
687. Yoakum, C. S. Psychology and Business. *Journal of Personnel Research*, 1922, 1, 306-318.
688. Yoakum, C. S. Standardized Methods of Employment. *Electric Railway Journal*, 1922, 59, 679-681.
689. Yoakum, C. S., and Bills, M. A. Tests for Office Occupations. *Annals of the American Academy of Political and Social Science*, 1923, 110, no. 199, 60-73.
690. Yoakum, C. S., and Yerkes, R. M. Army Mental Tests. New York, Holt, 1920, pp. 303.
691. Young, H. H. Intelligence Ratings and Success of Nurses in Training. *Journal of Applied Psychology*, 1924, 8, 377-389.
692. Yule, G. U. Introduction to the Theory of Statistics. London, Griffin, 1917, pp. 382.

## INDEX

- Academic record and vocational success, 372 ff.  
Acquaintance as affecting ratings, 360.  
Aerial observers, tests for, 207.  
Age  
  mental differences due to, 179 ff.  
  and turnover, 392.  
  and vocational aptitude, 390.  
"Alpha" test, 53, 91.  
Analysis of job preliminary to selecting tests, 190 ff., 226.  
Application blank, 386 ff.  
Application letter, 401 ff.  
Army Alpha test, 53, 91.  
Association, tests of, 72.  
Astrology, 17.  
Attention, tests of, 68.  
Attitude  
  of subjects, 114.  
  of workers and management, 497.  
Automobile drivers, tests for, 256, 458, 461.  
Automobile mechanics  
  occupational description, 479.  
  trade test for, 451.  
Average deviation, 160.  
Aviators, tests for, 54, 208, 245.
- Billers, tests for, 251.  
Binet, 49, 69.  
Blonde and brunette traits, 38.  
Bricklayers, trade test for, 452, 454.  
Bureau of Public Personnel Administration, 58, 496.
- Cabinet workers  
  trade test for, 451.  
  tests for, 251.  
Carpenters, trade test for, 453.  
Character analysis. *See* Physiognomy.  
Civil service, 495.  
Clerical workers  
  rating scale for, 343.  
  tests for, 246, 278.  
Clothing operators, intelligence of, 279.  
Cooks, trade test for, 452.  
Coöperative research, 492.  
Correction of ratings, 347.
- Correlation  
  interpretation of coefficient, 29, 213.  
  methods  
    products-moments, 199, 515.  
    rank-difference, 198, 511.  
    scatter plot, 200.  
    procedure, 29.  
    test with criterion, 198.  
Criteria, combining heterogeneous, 156 ff., 517.  
Criterion  
  availability of, 142, 174.  
  estimates by superiors, 143.  
  miscellaneous types, 154 ff.  
  need of, 141, 505.  
  production figures, 150 ff.  
Critical scores  
  in intelligence, 283.  
  in special capacity tests, 209, 221.  
  upper, 291.
- Decision, tests of, 78.  
Defined groups, rating by, 334, 422.  
Defining traits in rating scale, 326 ff.  
Designer in steel, occupational description, 481.  
Deviation  
  average, 160.  
  standard, 161, 164.  
Dial machine operators, tests for, 193 ff.  
Directions, tests of following, 80.
- Electrician  
  storage battery, trade test for, 453.  
  wireman, trade test for, 460.  
Emotion, tests of, 81.  
Employment psychology  
  fundamental principle, 5.  
  where valuable, 6.  
Engineers vs. salesmen, interests of, 308.  
Error, standard, 211.  
Estimates by superiors as criterion, 143.  
Executives  
  intelligence of, 279.  
  rating scale for, 332, 334, 341.  
Experience, previous, 173.
- Feather factory workers, tests for, 252.

- Firemen, tests for, 258.  
 Follow-up of tests, 260.  
 Fool-proof tests, 195.
- Gold bricks, 14 ff.  
 Graphic rating scale, 337 ff., 423.  
 Group tests, 103.  
 Guessing answer, probability of, 131.  
 Gun pointers, tests for, 207.
- Halo effect, 36, 358.  
 Height of salesmen, 388.  
 Hierarchy of occupations, 265 ff.  
 History of vocational psychology, 46 ff.  
 Hollerith machine operators, tests for, 251.  
 Honesty, tests of, 83.  
 Hospital attendants, tests for, 259.
- Incentive in tests, 126, 173.  
 Individual differences, interest in, 46.  
 Individual tests, 101.  
 Information test as measure of interest, 301, 313.  
 Ingenuity, tests of, 78.  
 Initial success in job prognostic of latter, 383.  
 Intelligence  
   of departments in an organisation, 273 ff.  
   kinds of, 87.  
   nature of, 86.  
   of occupational groups, 265 ff.  
   optimum vs. maximum, 285 ff.  
   and physiognomy, 33 ff.  
   quotient, 98.  
   and stability, 287.  
   surveys, 276.  
   tests of, 89 ff.  
   tests in job analysis, 485.  
   and vocational aptitude, 265 ff.
- Interest  
   and ability, 10, 296.  
   evaluation of measures of, 307.  
   measurement of, 298 ff.  
   permanence of, 295.  
   questionnaire, 298.
- Interview  
   employment, 414 ff.  
   giving information in, 424.  
   improvements in technique, 418.  
   for job analysis, 470.  
   unreliability of, 415.
- Interviewer  
   for job analysis, 471.
- Interviewer, rating scale for use of, 419.
- Janitors, tests for, 259.  
 Job analysis, 193, 467 ff.  
   interview, 470.  
   interviewer, 471.  
   need, 468.  
   purpose, 467.  
   work sheet, 475.
- Journalists, tests for, 260.
- Laboratory for employment tests, 196, 506.
- Lathe operators  
   tests for, 208.  
   trade tests for, 440.
- Learning, tests of, 70.
- Letter of application, 401 ff.
- Linear scale criterion, 146.
- Machinists, trade test for, 454.
- Maladjusted worker, social implications, 10, 507.
- Man-to-man rating scale, 329 ff., 420.
- Marital status of salesmen, 394.
- Mechanical aptitude, tests of, 94 ff.
- Mechanical interests vs. social, 312.
- Memory, tests of, 75.
- Mental components of occupation, tests for, 226 ff.
- Mental tests. *See* Tests.
- Motor control, tests of, 64.
- Motormen, tests for, 50, 205, 208, 255.
- Multiple correlation, coefficient of, 239 ff.
- National Research Council, 58, 495.
- Normal frequency curve, 162 ff.
- Objective score, 196.
- Objective traits vs. subjective, 356.
- Occupational description, 479. *See also*  
   Job analysis.
- Occupational groups, intelligence of, 265 ff.
- Office boys, intelligence of, 278.
- Officers' rating scale, 330.
- Omni-bus test, 96.
- Orientation of psychologist taking up  
   employment work, 188.
- Painters, trade test for, 451.
- Partial correlation, 55, 135, 233 ff.,  
   520 ff.
- Pattern makers, trade test for, 459.
- Percentile scores, 97, 266.



- Personal history blank, 386.  
 Personal history factors in vocational prediction, 386 ff., 397.  
 Personnel Research Federation, 58, 496.  
 Phrenology, 21 ff.  
 Physiognomy  
   commercial systems of, 27, 42.  
   experimental study of, 28 ff.  
   popular belief in, 25.  
 Policemen, tests for, 259.  
 Previous experience and vocational prediction, 395.  
 Probability of vocational success, prediction of, 210 ff.  
 Production figures as criterion, 150.  
 Psychological Corporation, 58, 496.  
 Public speakers, tests for, 260.  
  
 Quartile, 266.  
 Questionnaire  
   interest, 298.  
   job analysis, 470.  
  
 Railway postal clerks, trade test for, 457.  
 Ranking workers in criterion, 145.  
 Ranks, averaging, 517.  
 Rapport, 114, 186, 419.  
 Rating scale  
   interviewer's, 419.  
   in job analysis, 484.  
   methods  
     defined groups, 334 ff., 422.  
     graphic, 337 ff., 423.  
     man-to-man, 329 ff., 420.  
   officers', 56.  
   purpose, 317.  
 Ratings, correction of, 347.  
 Reaction time, tests of, 72.  
 Reasoning, tests of, 77.  
 Recommendation blank, 413.  
 Recommendations, 409 ff.  
 Regression equation, 204, 238 ff., 520 ff.  
 Reliability  
   of criterion, 149.  
   of letters of application, 402.  
   of rating scale, 345 ff.  
   of tests, 136, 197.  
  
 Salesmen  
   intelligence of, 274, 280.  
   interest tests for, 314.  
   tests for, 248, 457.  
   vs. engineers, interests of, 308.  
 Sampling of subjects, 176 ff.  
  
 School marks and vocational success, 372 ff.  
 Self-estimates, 408.  
 Sensory capacity, tests of, 67.  
 Sensory defects, 182.  
 Sexes, intelligence of, in office, 276.  
 Sex of subjects, 178.  
 Silk mill operatives, intelligence of, 282.  
 Social interests vs. mechanical, 312.  
 Space perception, tests of, 74.  
 Speed vs. accuracy score, 130.  
 Spiritualism, 18.  
 Standard deviation, 161, 164.  
 Standard error of estimate, 211.  
 Stenographers  
   tests for, 250.  
   trade tests for, 456.  
 Subjective traits vs. objective, 356.  
 Subjects used in evaluating tests, 172 ff.  
  
 Telegraphers, tests for, 208, 243, 254.  
 Telepathy, 19.  
 Telephone operators, tests for, 50, 208, 248, 253.  
 Temperament, tests of, 84.  
 Tests  
   administration of, 101 ff.  
   capacity vs. proficiency, 62, 431.  
   early development of, 48.  
   general nature of, 60.  
   general vs. special capacity, 62.  
   incentive, 126.  
   instructions, 120 ff.  
   in job analysis, 485.  
   limitation on response, 111 ff.  
   material  
     alternative, 119.  
     arrangement, 117.  
     difficulty, 115 ff.  
   method of response  
     oral, 106.  
     performance, 108.  
   proficiency vs. capacity, 62, 431.  
   rapport, 114.  
   reliability, 136.  
   scoring, 127 ff.  
   sensitivity, 119.  
   special vs. general capacity, 62.  
   time limit, 111.  
   type of response, 109 ff.  
   work limit, 112.  
   written method, 107.  
 Time clerks, trade test for, 455.  
 Tire finishers, tests for, 227.  
 Total mental situation, tests for, 194 ff.

Trade test, 55, 431 ff.  
 calibration, 446.  
 in job analysis, 485.  
 kinds, 434.  
 methods of developing, 438 ff.  
 need, 431.  
 oral, 434, 450.  
 performance, 435, 459.  
 picture, 434, 453.  
 written, 435, 454.  
 Training of raters, 362.  
 Traits  
   defining in rating scale, 326 ff.  
   selection of for rating scale, 320 ff.  
 Typesetters, tests for, 253.  
 Typists, tests for, 249.

Validity  
   of letters of application, 405.  
   of ratings, 353.  
   of tests, 198.  
 Vocational guidance vs. vocational selection, 8.  
 War, psychology in the, 51 ff.  
 Weavers, tests for, 252.  
 Weighting tests, 55, 231.  
 Weighting traits in rating scale, 323.  
 Wireman, trade test for, 460.  
 Work sheet for job analysis, 475.  
 Worker in his work unit, 9.